DOCUMENT RESUME

ED 455 642		EC 308 530

AUTHOR	Miller, Phyllis, Ed.
TITLE	What Does Being a G Wiz Mean in Real Life?
INSTITUTION	American Mensa Education and Research Foundation, Arlington, TX.
PUB DATE	2001-00-00
NOTE	97p.; Theme issue. Published three times a year.
AVAILABLE FROM	Mensa Education and Research Foundation, 1229 Corporate Dr. West, Arlington, TX 76006-6103. Tel: 973-655-4225; Fax: 973-655-7382; e-mail: millerp@mail.montclair.edu.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Mensa Research Journal; v32 n2 2001
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Admission (School); Adults; *Aging (Individuals); *Gifted; *Graduate Study; *Higher Education; *Intelligence Tests; Knowledge Level; *Life Satisfaction; Longitudinal Studies; Older Adults; *Well Being
IDENTIFIERS	Graduate Record Examinations

ABSTRACT

The articles in this journal issue examine three different aspects of gifted adulthood. In "Self-Appraisal, Life Satisfaction, and Retrospective Life Choices across One and Three Decades" (Carole K. Holahan and others), 383 children who were part of Terman's original study of the gifted in 1921 were revisited at various points in their lives. Participants who reported living up to their intellectual abilities were higher in overall life satisfaction and were less likely to report that they would make different choices in work or family life three decades later. "Consequences of How We Define and Assess Intelligence" (Wendy J. Williams), considers some of the consequences of how intelligence is defined and assessed in young adults. In particular, it discusses the Graduate Record Examination and its usefulness in predicting success in graduate school. The use of intelligence tests in general is based on a certain definition of intelligence, and the article argues that such a definition is not necessarily what is needed to determine success in school. The last article, "The Locus of Adult Intelligence: Knowledge, Abilities, and Nonability Traits" (Phillip L. Ackerman and Eric L. Rolfhus), draws the distinction between general intelligence and knowledge, and studies the relationship of both to the aging process. (Articles include references.) (CR)

What Does Being a G Wiz Mean in Real Life?

Phyllis Miller, Editor

Mensa Research Journal
Volume 32, Number 2

Published by Mensa Education & Research Foundation and Mensa International, Ltd.™

MENSA Research Journal

What does being a *g* wiz mean in real life?

Vol. 32, No. 2

MENSA
Research Journal

Vol. 32, No. 2

Table of Contents

Mensa Research Journal is published three times a year, in conjunction with Mensa International, Ltd., by the Mensa Education & Research Foundation, Dr. Michael Jacobson, president, 1840 N. Oak Park Ave., Chicago, IL 60635-3314.

Staff:

Editor • Phyllis Miller, 23 Lexington Road, Somerset, NJ 08873, MRJ@merf.us.mensa.org

Associate Editor • Francis Cartier, 1029 Forest Ave., Pacific Grove, CA 93950, fcar889755@cs.com

Circulation Manager • Patty Wood, American Mensa, Ltd., 1229 Corporate Drive West, Arlington, TX 76006-6103, PattyW@AmericanMensa.org.

Production Manager • Annette L. Kovac, American Mensa, Ltd.,1229 Corporate Drive West, Arlington, TX 76006-6103, AnnetteK@AmericanMensa.org.

The *Journal* is published three times a year. Subscribers will receive all issues for which they have paid, even if frequency of publication varies. Membership in Mensa is not required. To subscribe, order back issues, or report address changes, write to *Mensa Research Journal*, 1229 Corporate Drive West, Arlington, TX 76006-6103.

The Mensa Education & Research Foundation

Mensa, the high IQ society, provides a meeting of the minds for people who score in the top 2 percent on standardized IQ tests. As an international organization with thousands of members worldwide, Mensa seeks to identify and foster human intelligence; encourage research in the nature, characteristics and uses of intelligence; and provide a stimulating intellectual and social environment for its members.

The first two of these purposes are largely carried out by the Mensa Education & Research Foundation (MERF). MERF is a philanthropic, nonprofit, tax-exempt organization funded primarily by gifts from Mensa members and others. MERF awards scholarships and research prizes, publishes the *Mensa Research Journal,* and funds other projects consistent with its mission.

For more information about the Mensa Education & Research Foundation, write to MERF, 1229 Corporate Drive West, Arlington, TX 76006-6103. For information about joining American Mensa, call 1-800-66-MENSA. See also www.us.mensa.org. If you reside outside the U.S., see the Mensa International, Ltd., Web site at www.mensa.org.

To renew your subscription or subscribe to the *Mensa Research Journal*, just fill out the form on the next page (or a photocopy of it) and mail it with a check in the appropriate amount. Please refer to your mailing label now to determine how many issues of the *MRJ* you will receive before your subscription expires.

To: MERF, 1229 Corporate Drive West, Arlington, Texas 76006-6103

Please enter my subscription to the *Mensa Research Journal*.

❒ New ❒ 3 issues U.S. $21 (outside U.S. $25)

❒ Renewal ❒ 6 issues U.S. $42 (outside U.S. $50)

❒ Sample *Mensa Research Journal* U.S. $7 (outside U.S. $9)

Subscription amount enclosed $ ____________________

Please make payments in U.S. funds.

I would like to make the following tax-deductible contribution to MERF.

❒ Contributor ($25 to $99)	$ ________	Allocation (if desired)
❒ MERF Donor ($100-249)	$ ________	❒ General support
❒ Bronze Donor ($250-499)	$ ________	❒ Scholarships
❒ Silver Donor ($500-999)	$ ________	❒ Gifted Children Programs
❒ Gold Donor ($1,000-2,499)	$ ________	❒ Addition to endowed fund
❒ Platinum Donor ($2,500+)	$ ________	named ____________

I would like my contribution to be a ❒ Memorial honoring ____________________

❒ Tribute honoring ____________________

Please send acknowledgement or notice to ____________________

Address ____________________

City/State/Zip ____________________

Name and address of contributor/subscriber:

Name ____________________ Member # ____________

Address ____________________

City/State/Zip ____________________ MRJ

For Office Use

Date received: ________ Total $ ________ Bank: ____________ Date entered: ________

Sub code: ________ First issue number: ____________ Date sent: ________

Editor's Preface

A great deal of attention is given to gifted children these days, and rightly so. As the twig is bent, etc., etc., and those of us who have gifted children or were gifted kids ourselves know that gifts can be stunted if they are not properly nurtured. However, much less attention is given to what happens when we grow up. The three articles in this issue of the *Mensa Research Journal* examine three different aspects of gifted adulthood.

Do you think you have lived up to your intellectual abilities? Do you think maybe you should have chosen a more intellectually demanding profession? At the age of 80, when you look back on your life, will you be satisfied? These are the questions Carole Holahan, Charles Holahan, and Nancy Wonacott sought to answer when they revisited the children who were part of Terman's original study of the gifted. Lewis Terman began a study of a group of gifted children in 1921, and the Terman Study of the Gifted is now the oldest and most complete study of the human life cycle. These gifted people have been interviewed at various points in their lives, and their thoughts about how they have lived and the choices they made is fascinating reading.

Wendy Williams considers some of the consequences of how intelligence is defined and assessed in young adults. In particular, she discusses the Graduate Record Examination (GRE), which is used extensively in this country in the graduate school admissions process, and its usefulness in predicting success in graduate school. The use of intelligence tests in general is based on a certain definition of intelligence, and Williams argues that such a definition is not necessarily what is needed to determine success in school.

In the final piece, Phillip Ackerman and Eric Rolfhus draw the distinction between general intelligence and knowledge, and study the relationship of both to the aging process. Why is it that when you play Trivial Pursuit™ with your kids, you always win? They may be intelligent, more intelligent than you, but you have greater knowledge. I'm always amazed at my children's lack of knowledge — they are in their 30s, and there is so much they don't know. It occurs to me that when I was in my 30s, my parents felt the same way about me!

Phyllis Miller
Editor

8

Self-Appraisal, Life Satisfaction, and Retrospective Life Choices Across One and Three Decades

Carole K. Holahan, Charles J. Holahan, and Nancy L. Wonacott,
University of Texas at Austin

This research investigated the relationship of a self-appraisal of having lived up to one's intellectual abilities at midlife (average age of 49 years) with life satisfaction and retrospective life choices one and three decades later among 383 participants in the Terman Study of the Gifted. Study 1 showed that participants who reported living up to their intellectual abilities were higher in satisfaction with occupational success, satisfaction with family life, and joy in living 11 years later. Study 2 showed that participants who reported living up to their abilities were higher in overall life satisfaction and were less likely to report that they would make different life choices in work or family life three decades later. In an integrative structural equation model, the relation between the midlife self-appraisal of having lived up to intellectual abilities and overall satisfaction at age 80 was mediated by life satisfaction discrepancy at age 61.

Self-concept theorists increasingly view the self as comprised of a variety of representations (Markus & Wurf. 1987). Because the self-concept conveys representations of one's actual and ideal self (Higgins, 1987), it can shape affective and motivational outcomes in a powerful and enduring way (Ross & Conway, 1986). The present research examined a component of the self-concept likely to play a key role in adult development — a midlife appraisal of having lived up to one's intellectual abilities — among members of the Terman Study of the Gifted (Terman, 1925). The research examined several aspects of life satisfaction approximately one and three decades later and explored the relation of later life satisfaction with alternative life choices.

One aspect of the self-concept, that which reflects the "self that might have been," is a topic of emerging interest (Landman & Martis, 1992; Landman, Vandewater, Stewart, & Malley, 1995). Markus and Wurf (1987) theorized that the self one would like to be provides a conceptual anchor for evaluating one's current self. Similarly, Higgins and colleagues (Higgins, 1987; Higgins, Bond, Klein, & Strauman, 1986) have emphasized that discrepancies between one's actual and one's ideal self relate to disappointment and dissatisfaction.

Markus and Wurf (1987) also proposed that a central feature of the self-concept is its motivational function. For example, they theorized that the self one would like to be operates as an incentive. In a similar vein, Markus and Nurius

Correspondence concerning this article should be addressed to Carole K. Holahan, Department of Kinesiology and Health Education, University of Texas at Austin. Bellmont Hall 222 (D3700), Austin, TX 78712. Electronic mail may be sent to c.holahan@mail.utexas.edu.
Reprinted from *Psychology and Aging,* Vol. 14, No. 2, 1999.

(1986) proposed that one's possible selves may be construed as "the cognitive component" of motivation (p. 954). Higgins (1987) likewise suggested that frustration from unfulfilled desires underlies the motivational aspect of the actual-ideal self discrepancy. The negative emotion cued by life regrets, in turn, motivates further efforts to cognitively undo aversive events (Roese & Olson. 1995).

Individuals' reflections on the self that might have been may be especially salient at the midlife transition. For example, Helson (1992) has found midlife to be an important time for the revision of women's self-conceptualizations. Levinson's (1978, 1996) theories of the male and female life cycles also emphasize midlife as a critical time for reassessment. Moreover, the consequences of life regrets may be especially apparent in the aging years. In a study of older persons, Erikson and his colleagues (Erikson, Erikson, & Kivnick, 1986) found that many of the individuals in their study were engaged in a positive reassessment of their earlier life in aging, which enabled them to successfully balance feelings of ego integrity and a sense of despair.

The purpose of the present research was to investigate the relationship of a midlife appraisal of having lived up to one's intellectual abilities with (a) life satisfaction assessed approximately one decade later and (b) both life satisfaction and life choices individuals would make differently assessed three decades later. Participants were 188 men and 195 women in the Terman Study of the Gifted (Terman, 1925). The midlife appraisal was made when participants, who were an average of 49 years of age, answered a question asking whether they had lived up to their intellectual abilities. Life satisfaction in several areas was indexed at an average age of 61. In addition, overall life satisfaction and life choices that participants would make differently if they could live their lives again were measured at an average age of 80.

The Terman Study of the Gifted is the oldest and most complete study of the human life cycle (see Holahan, 1984, 1988; Holahan & Sears, 1995; Oden, 1968; P. S. Sears & Barbee, 1977; R. R. Sears, 1977; Terman, 1925). It was begun in 1921 by Lewis Terman and eventually included 1,528 gifted children. The average age of the core sample selected in 1921 was 11. The study has been extended over the ensuing decades, with the latest data collection in 1996.

The issue of living up to one's intellectual abilities is particularly relevant to a sample of individuals selected for their high intelligence. The criterion for selection was a minimum IQ of 135. The individuals in the Terman sample achieved high levels of education relative to others at that time, with about 76 percent of men and 70 percent of women graduating from college, compared with eight percent in the general population (Holahan & Sears, 1995). Over the years, the data collections in the Terman study have emphasized achievement. They have routinely requested information on education and career pursuit, as well as honors received in occupational, community, or other contexts.

Study 1

From the Terman study archival data, there was an opportunity to examine the participants' midlife assessments of their having lived up to their intellectual abilities as reported in 1960. The archives also afforded an opportunity to view the affective correlates of this appraisal. In 1972, the study participants were asked to evaluate their life satisfaction in the occupational and family spheres, as well as their joy in living. The areas of work and family were chosen for analysis in the present study because of their central roles in theories concerning the development of a life structure (e.g., Erikson et al., 1986; Levinson, 1978).

Based on conceptualizations of the affective correlates of the "self that might have been" (Higgins, 1987; Higgins et al., 1986; Markus & Wurf, 1987), we hypothesized that individuals who reported in midlife that they had not lived up to their intellectual abilities would score lower on all three of these indexes of life satisfaction 12 years later. Based on traditional gender roles and the differential occupational opportunities open to the Terman men and women (Holahan & Sears, 1995), it was also expected that the subjective assessment of having lived up to one's intellectual potential would be more tied to satisfaction with occupational success for the men than for the women. In addition, reasoning that life regrets would predict subsequent perceived goal-related discrepancies, we hypothesized that individuals who reported in midlife that they had not lived up to their intellectual abilities would score lower on indexes of life satisfaction discrepancy (i.e., satisfaction adjusted for goal importance) pertaining to occupational and family life and joy in living 12 years later. Moreover, we predicted that these relations would be independent of prior mental health and objective achievement and would hold controlling for these variables.

Method

Participants

Some overall selection criteria pertained to both Studies 1 and 2. Participants in both studies were members of the Terman Study of the Gifted who responded to a question tapping the self-appraisal of having lived up to their intellectual abilities in the 1960 survey when they were an average of 49 years of age. To ensure a more homogenous age sample and consistency across the two studies, participation in both studies was restricted to individuals who were at least 75 years of age at the 1992 survey and who had responded in 1972. The age restriction excluded 42 younger members of the Terman Study from the present analyses. These procedures resulted in highly comparable samples across the two studies. Among individuals who responded to the 1992 survey, the focus of Study 2, 91 percent had also responded to the 1972 survey, the focus of Study 1.

It should be noted that attrition has made the sample more select in some

areas. Participants who have remained in the study are similar to those who left the study in terms of IQ and socioeconomic status of family of origin. However, the continuing sample has more education, better health, and greater occupational success (men only) than those who left the study (Holahan & Sears, 1995). It would seem, however, that by restricting the range in the sample such attrition would have made the analyses reported here more conservative.

The maximum number of participants for whom data were available for the present analyses was 383 (188 men and 195 women). The number of participants in some analyses was less than 383 due to missing values on some variables.

The return rate for the 1972 survey was 7,596. At the 1972 survey, participants were a mean age of 61.

Measures

Appraisal of having lived-up. In 1960, the participants were asked, "On the whole, how well do you think you have lived up to your intellectual abilities? Don't limit your answer to economic or vocational success only." Response options varied along a 5-point scale, ranging from 1 (Consider my life largely a failure) to 5 (Fully). Responses were coded into two categories: not lived up versus lived up. The not lived-up category included responses of "consider my life largely a failure,"[1] "far short," and "considerably short." The lived-up category included "reasonably well" and "fully."

Life satisfaction. In 1972, the participants were asked to rate their satisfaction with their life experience in several domains. Three life satisfaction domains were analyzed in the present study: occupational success, family life, and joy in living. In each area, participants were asked to check one of the following five response alternatives: 1 = found little satisfaction in this area: 2 = on the whole, somewhat dissatisfied; 3 = had a mixed experience but am not discontented; 4 = had a satisfactory degree of success; and 5 = had excellent fortune in this respect. (For examples of research using these items, see Holahan & Sears, 1995; P. S. Sears & Barbee, 1977; R. R. Sears, 1977.)

Life satisfaction discrepancy. In 1972, the participants were also asked to rate the importance of their life goals in the plans they made for themselves in early adulthood in each of the life satisfaction domains of occupational success, family life, and joy in living. In each domain, participants were asked to check one of the following four response alternatives: 1 = less important to me than to most people, 2 = looked forward to a normal amount of success in this respect, 3 = expected a good deal of myself in this respect, and 4 = of prime importance to me, was prepared to sacrifice other things for this. A life satisfaction discrepancy score in each of the three domains was computed by subtracting the importance of each domain from its satisfaction score. The discrepancy scores had a

---

[1] Only 1 participant chose the response option "consider my life largely a failure."

range of -3 to 4, with lower scores representing less satisfaction relative to goal importance.

Occupational level. Participants' occupations were rated according to Duncan's socioeconomic index (Miller & Miller, 1977). Occupations listed in 1972 were used primarily in assigning occupational ratings. Where necessary, information from other surveys (1960 or 1977) was used to substantiate or clarify ratings. In the present research, occupations were coded into one of three levels: 1 = lower-level occupations (Duncan scores of 0-610), 2 = administrative and minor professional (Duncan scores of 617-771), and 3 = major professional (Duncan scores of 774-960). Homemakers were classified in the first level.

Mental health. In 1960, cumulative ratings of mental health were made for each Terman study participant (see Oden, 1968). All ratings utilized multiple sources of information from each follow-up survey since 1940, such as "personal conferences with the subject or members of his family by the research staff, responses by the subjects to questionnaire inquiry, reports by parents and spouses of the subjects, and letters or other personal communications from the subjects or other qualified informants" (Oden, 1968, p. 8). Based on this information, each participant's mental health was coded into one of three levels, with higher scores reflecting poorer adjustment: 0 = satisfactory adjustment (e.g., only minor and realistic anxieties), 1 = some difficulty in adjustment (e.g., psychiatric or other help sought), and 2 = serious difficulty in adjustment (e.g., interference with marriage, occupation, or social relationships or psychiatric hospitalization). (For a recent application of these mental health data, see Martin et al., 1995.)

Results

Predictors of Lived Up

Initially, we examined three variables that might be predictively related to the lived-up self-appraisal: the three-level Duncan socioeconomic index, gender, and the three-level 1960 cumulative measure of mental health. A 2 x 3 x 2 x 3 (Lived Up x Occupational Level x Gender x Mental Health) hierarchical log linear analysis contained the four main effects and the following pairwise interactions: Lived Up x Occupational Level, Occupational Level x Gender, Lived Up x Gender, and Lived Up x Mental Health. The goodness of fit for the model was satisfactory, $G^2(22) = 28.35$, $p = .164$, $N = 383$.

Follow-up chi-square analyses indicated that a greater proportion of individuals in higher-level occupations responded that they had lived up to their intellectual abilities, $c^2(2, N = 383) = 8.88$, $p < .05$. In addition, a greater proportion of men than women said that they had lived up to their intellectual abilities (70.2 percent of men as compared with 59.5 percent of women), $c^2(1, N = 383) = 4.82$, $p < .05$. Finally, the proportion of lived-up responses was positively associated with the mental health rating, $c^2(2, N = 383) = 11.26$, $p < .01$.

Life Satisfaction in 1972

Participants' satisfaction with their experience in several domains as reported in 1972 was analyzed as a function of their report in 1960 of having lived up to their intellectual abilities and by gender. A 2 x 2 (Lived Up x Gender) multivariate analysis of covariance (MANCOVA) was run with satisfaction with occupational success, family life, and joy in living as the dependent variables, and occupational level and mental health as covariates. The MANCOVA was significant for lived up (Wilks lambda = .90), $F(3, 296) = 10.37$, $p < .001$. There was not a significant multivariate effect for gender or for the Lived Up x Gender interaction. In follow-up univariate analyses of covariance (ANCOVAs), there was a significant lived-up effect for satisfaction with occupational success, $F(1, 311) = 26.89$, MSE = .86, $p < .001$; satisfaction with family life, $F(1, 333) = 4.87$, MSE = .92, $p < .05$; and joy in living, $F(1, 332) = 13.85$, MSE = .82, $p < .001$, with means higher for the group reporting having lived up to their abilities. Table 1 presents the means on the three variables for men and women separately.

In the univariate ANCOVAs, there was a significant gender effect only for satisfaction with occupational success, $F(l, 311) = 5.04$, MSE = .86, $p < .0$, with men reporting higher satisfaction. In addition, there was a significant Lived Up x Gender interaction only for satisfaction with occupational success, $F(1, 311) = 5.27$, MSE = .86, $p < .05$, with men who reported they had lived up to their abilities particularly satisfied with their occupational success. Post hoc *t*-tests conducted within gender groups demonstrated that the lived-up effect for satisfac-

Table 1

Mean Satisfaction With Occupational Success, Family Life, and Joy in Living as Reported in 1972 for Men and Women Who Reported in 1960 That They Had or Had Not Lived Up to Their Intellectual Abilities

	Not lived up			Lived up		
Domain	**M**	**SD**	**n**	**M**	**SD**	**n**
			Men			
Occupational success	3.40	0.97	50	4.25	0.74	119
Family life	3.92	1.03	50	4.33	0.93	118
Joy in living	3.71	0.89	49	4.15	0.85	118
			Women			
Occupational success	3.26	1.08	57	3.63	1.05	91
Family life	4.00	1.14	67	4.22	0.91	104
Joy in living	3.76	1.22	66	4.24	0.82	105

tion with occupational success was significant for both gender groups, with the effect stronger for men, $t(167) = 6.22, p < .001$, than for women, $t(146) = 2.03$, $p < .05$.

Life Satisfaction Discrepancy in 1972

Life satisfaction discrepancy (i.e., life satisfaction-goal importance) was analyzed in a 2 x 2 (Lived Up x Gender) MANCOVA. Life satisfaction discrepancy pertaining to occupational success, family life, and joy in living were dependent variables, and occupational level and mental health were covariates. The MANCOVA was significant for lived up (Wilks lambda = .95), $F(3, 272) = 4.65, p < .01$, and for gender (Wilks lambda = .95), $F(3, 272) = 3.96, p < .01$, but not for the Lived Up x Gender interaction. The group that reported having lived up to their intellectual abilities showed more favorable scores (i.e., less discrepancy in the direction of negative self-assessment) than the not lived-up group, and men showed more favorable scores than did women.

In follow-up univatiate ANCOVAs, there was a significant lived-up effect for life satisfaction discrepancy, with occupational success, $F(1, 297) = 5.23$, MSE = 1.24, $p < .05$; family life, $F(1, 322) = 5.70$, MSE = 1.06, $p < .05$; and joy in living, $F(1, 310) = 10.65$, MSE = .84, $p < .01$. The group that reported having lived up to their abilities had more favorable scores on all three variables. There was a significant gender effect for life satisfaction discrepancy only with family life, $F(1, 322) = 6.82$, MSE = 1.06, $p < .01$, with men showing more favorable scores than women. There were no significant univariate effects for the Lived Up x Gender interaction.

Study 2

The Terman study archives also afforded an opportunity to view longer-term affective and motivational correlates of the participants' 1960 assessment of their appraisal of having lived up to their intellectual abilities. In 1992, study participants were asked about their overall life satisfaction and about what choices they would make differently if they could live their lives again.

Based on conceptualizations of the long-term affective correlates of the self that might have been (Higgins, 1987; Markus & Wurf, 1987), we hypothesized that individuals who reported that they had not lived up to their intellectual abilities in midlife would score lower on overall life satisfaction 30 years later. Moreover, based on the view that the negative emotion cued by a negative comparison with what might have been motivates efforts to cognitively undo the aversive event (Roese & Olson, 1995), we hypothesized that individuals who reported that they would make different choices in either the work or family domains would report lower levels of life satisfaction than those who would not make any choices differently. Moreover, we predicted that these relations would

be independent of prior mental health, objective achievement, and general health in 1992 and would hold controlling for these variables.

Based on conceptualizations of the motivational correlates of the self that might have been (Higgins, 1987; Markus & Nurius, 1986; Markus & Wurf, 1987), we hypothesized that the tendency to make different choices in the work or family domains in contrast to the tendency to change nothing would be predicted by the midlife assessment of having lived up to intellectual abilities 30 years earlier. Finally, we tested an integrative model of the associations among the 1960 lived-up variable, life satisfaction discrepancies in 1972, and overall satisfaction in 1992 in a structural equation model (SEM) using LISREL 8 (Joreskog & Sorbom, 1993). Reasoning that an earlier self-appraisal would operate through subsequent self-referent thought in predicting future outcomes, we hypothesized that the relationship between the 1960 self-appraisal of having lived up to intellectual abilities and overall satisfaction in 1992 would be mediated by life satisfaction discrepancy in 1972.

Method

Participants

Participants in Study 2 were members of the Terman Study of the Gifted who responded to the follow-up survey in 1992 and who also met the overall selection criteria described in Study 1. The return rate for the 1992 survey was 769. At the 1993 survey, participants ranged in age from 75 to 88 years, with a mean age of 80. Due to missing data, the maximum sample size in Study 2 analyses was 365 (178 men and 187 women).

Measures

Overall life satisfaction. In 1992, participants were asked, "All things considered, how satisfied are you with your life these days? "[2] Response options varied along a 9-point scale, ranging from 1 (completely dissatisfied) to 9 (completely satisfied). Single items indexing global life satisfaction have been used extensively in survey research and have acceptable psychometric characteristics (see Campbell, Converse, & Ropers, 1976; Sauer & Warland, 1982).

Alternative life choices. In 1992, the participants were asked in an open-ended question: "Looking back over your whole life what choices would you make differently?" Responses had been content coded earlier by Terman study research staff, who were blind to participants' appraisal of having lived up

[2]The correlation of overall life satisfaction in 1992 with the three satisfaction scores in 1972 was low to moderate (satisfaction with occupation, family life, and joy in living was .32, .18, and .34, respectively), making stability of life satisfaction less plausible as an alternative explanation for the study findings.

to their intellectual abilities in 1960 and to the present hypotheses. Consistent with the present emphasis on the work and family domains, responses for analysis were selected from three content categories: no change, family, and work. The no change category included responses such as "no changes," "no regrets," and "quite satisfied with choices." The work category included responses such as "chose wrong occupation," "would have liked a different career," and "should have aimed higher in career." The family category included responses such as "would have chosen different mate," "might have tried harder to be married," and "would spend more time in family relationships."

General health. In 1992, participants were asked a question concerning their general health since 1986. Response options varied along a 5-point scale, ranging from 1 (very poor) to 5 (very good). A two-level (good vs. poorer) health variable was defined as follows: Individuals who reported "good" or "very good" health (69.89 percent) were included in a good health group; individuals who reported "very poor," "poor," or "fair" health (30.2 percent) were included in a poorer health group. Self-ratings of general health have good construct validity and tend to be positively correlated with physicians' ratings (LaRue, Bank, Jarvik, & Hetland, 1979). Moreover, such ratings predict mortality beyond predictions based on objective indicators, such as physicians' assessments from physical examinations (Idler & Karl, 1991).

Results

Overall Satisfaction in 1992

The relationship of the 1960 lived-up variable with 1992 overall satisfaction was analyzed in a 2 x 2 x 2 (Lived Up x Health x Gender) ANCOVA. Occupational level and 1960 mental health were used as covariates. The 1992 measure of overall life satisfaction was the dependent variable. The analysis was significant for lived up, $F(1, 355) = 10.71$, $MSE = 2.25$, $p < .001$, and for health, F(1, 355> = 27.64, $MSE = 2.25$, $p < .001$. The Lived Up x Health interaction was nonsignificant. For the lived-up factor, the mean of the group reporting having lived up to their intellectual abilities was higher than that of the not lived-up group *Ms* = 6.95 and 6.30, respectively). For health, the satisfaction of participants reporting good health was higher than that of participants reporting poorer health (*Ms* = 6.98 and 5.99, respectively).

Alternative Life Choices

To investigate the relation of appraisal of having lived up to intellectual abilities in 1960 with alternative life choices as reported in 1992, a 2 x 3 x 2 (Lived Up x Choice x Gender) hierarchical log linear analysis was run as a saturated

Table 2

Distribution of Alternative Life Choices Reported in 1992 for Men and Women Who Reported in 1960 That They Had of Had Not Lived Up to Their Intellectual Abilities

	Men		Women	
Choice	**n**	**%**	**n**	**%**
		Not lived up		
No change	10	27.8	23	53.5
Work	19	52.8	9	20.9
Family	7	19.4	11	25.6
		Lived up		
No change	47	72.3	38	56.7
Work	10	15.4	11	16.4
Family	8	12.3	18	26.9

model. The three choice categories selected for analysis were no change, alternative choices in the work domain, and alternative choices in the family domain. Both the Lived Up x Choice interaction, $c^2(2, N = 211) = 13.09$, $p < .01$, and the Choice x Gender interaction, $c^2(2, N = 211) = 6.31$, $p < .05$, were significant. In addition, the three-way interaction (Lived Up x Choice x Gender) was signif icant, $c^2(2, N = 211) = 7.80$, $p < .05$.[3] Table 2 gives the distribution of alternative choices across the three choice categories by gender across levels of the lived-up variable.

Follow-up chi-square analyses within gender groups indicated that men who felt they had not lived up to their intellectual abilities, compared with men who felt they had lived up to their abilities, were more likely to say they would make life choices differently, $c^2(2, N = 101) = 20.22$, $p < .001$. The predominant response of men who did not live up to their abilities was to alter life choices in the work domain. For women, in contrast, the responses of those who did and those who did not live up to their abilities were comparably distributed across the no change, work, and family categories, $c^2(2, N = 110) = .36$, ns.

Relation of Satisfaction to Alternative Life Choices

The relation between alternative life choices and life satisfaction was investigated in a 2 x 2 x 2 (Choice x Health x Gender) ANCOVA, with occupational

---

[3] A small number of participants ($n = 11$) gave responses in both the work and family categories. An additional log linear analysis was run with the responses of these individuals included in a choice category. The results were essentially the same as those for the three-level choice category as reported above.

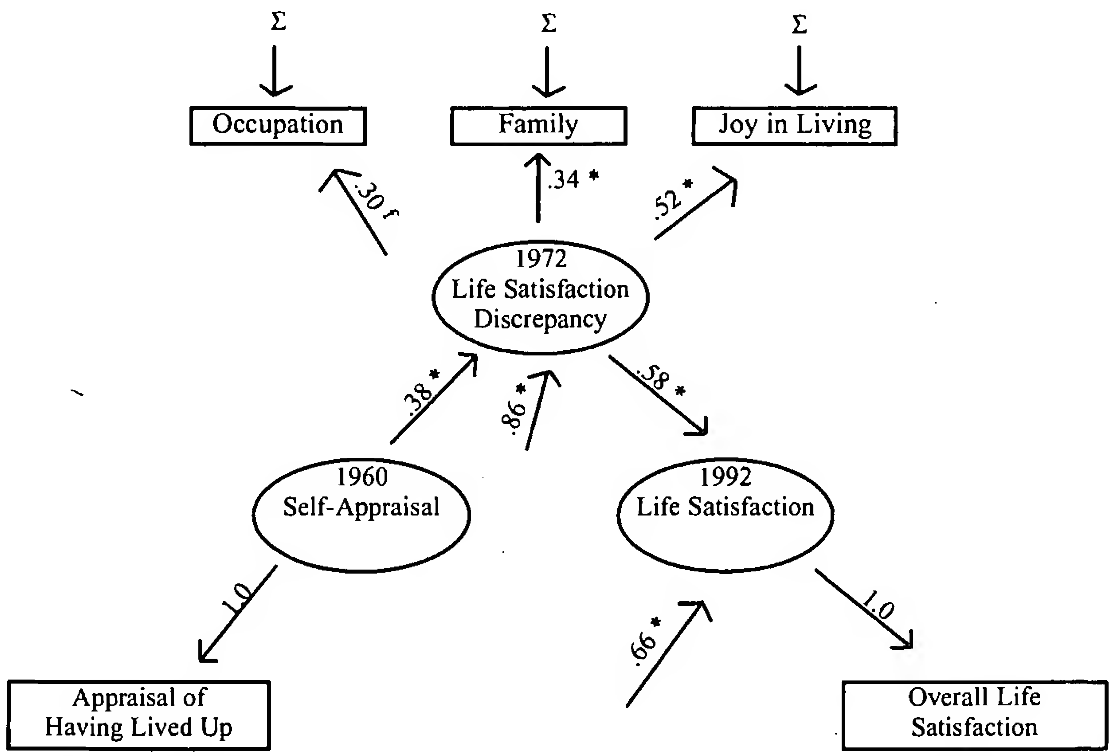
level and 1960 mental health as covariates. The choice factor was defined as stating no changes would be made versus stating that different choices would be made pertaining to either work or family. The 1992 measure of overall life satisfaction was the dependent variable. The results were significant for choice, $F(1, 204) = 17.26$, $MSE = 2.10$, $p < .001$, and health, $F(1, 204) = 15.41$, $MSE = 2.10$, $p < .001$. The no-change group reported higher satisfaction than the change group (*Ms* = 7.29 and 6.33, respectively). In addition, the good-health group reported higher satisfaction than the poorer-health group (*Ms* = 7.06 and 5.96, respectively).

An Integrative Longitudinal Model

We tested an integrative longitudinal model of the associations among the 1960 lived-up variable, life satisfaction discrepancies in 1972, and overall satisfaction in 1992 in a latent variable SEM using LISREL 8 (Joreskog & Sorbom, 1993). The 1960 appraisal of having lived up to intellectual abilities (coded dichotomously as "not lived up" = 0, "lived up" = 1) was an exogenous variable, and overall satisfaction in 1992 was an outcome variable (both measured with single indicators). Life satisfaction discrepancy in 1972 (measured with three indicators — life satisfaction discrepancy pertaining to occupational success, family life, and joy in living) was included as a mediating variable between the 1960 self-appraisal and 1992 satisfaction. To provide a metric for the latent constructs and to identify the measurement model, the first indicator loading for each latent construct was set to 1.0 in the unstandardized solution for the model. Variance-covariance matrices were used in the LISREL analyses.

The results of the LISREL test of the hypothesized model are presented graphically in Figure 1. The model provides a good fit to the data, overall $C^2(4, N = 313) = 2.55$, $p > .60$; adjusted GFI = .99. Based on examination of the modification indices, a parameter reflecting correlation between the unique variances for the measures of life satisfaction discrepancy pertaining to family life and joy in living was included in the model. All parameter estimates for the measurement model of the life satisfaction discrepancy latent construct and all parameter estimates in the structural model are significant at the .01 level. As predicted, the relationship between the 1960 self-appraisal of having lived up to intellectual abilities and overall satisfaction in 1992 was mediated by the life satisfaction discrepancy in 1972. The simple correlation in the model between the 1960 lived-up appraisal and 1992 overall satisfaction was significant, $r = .22$, $p < .01$. However, consistent with the mediational interpretation, when a direct path between the 1960 lived-up variable and 1992 overall satisfaction is added to the model, model fit is not significantly improved, $C^2(1, N = 313) = 2.25$, $p > .10$.

Figure 1



Results of the LISREL test (standardized estimates) of the structural equation and measurement models for an integrative longitudinal model. Latent constructs are shown in ellipses, and observed variables are shown in rectangles. f indicates a parameter set to 1.0 in the unstandardized solution: Σ represents unique variance in the three indicators of life satisfaction discrepancy. * $p < .01$.

General Discussion

Consistent with conceptualizations of the affective correlates of the self that might have been (Higgins, 1987; Higgins et al., 1986; Markus & Wurf, 1987), we found in Study 1 that a self-appraisal of having lived up to one's intellectual potential in midlife predicted satisfaction in the work and family domains and joy in living 12 years later. In Study 2, we found that the same midlife appraisal also predicted overall life satisfaction three decades later.

Specifically, individuals who reported that they had lived up to their intellectual abilities were more satisfied than were individuals who reported that they had not lived up to their abilities in each of the life domains assessed 12 years later and in overall satisfaction three decades later. Moreover, in an integrative structural equation model, we showed that the relation between a midlife self-appraisal of having lived up to intellectual abilities and overall satisfaction at age 80 was mediated by satisfaction discrepancy at age 61.

Consistent with conceptualizations of the motivational correlates of the self

that might have been (Higgins, 1987; Markus & Nurius, 1986; Markus & Wurf, 1987), we found in Study 2 that individuals' self-appraisals of having lived up to their intellectual abilities in midlife were related significantly to life choices they would make differently as reported three decades later. Individuals who reported that they had lived up to their intellectual abilities were more likely to say that they would not make any life choices differently. In contrast, individuals who reported that they had not lived up to their intellectual abilities were more likely to say that they would make different life choices in the work or family domains.

In Study 2, we also found that the life choices individuals would make differently were related significantly to overall life satisfaction. Although these correlational findings do not demonstrate direction of effect, they are consistent with the view that the negative emotion cued by an initial negative comparison motivates further efforts to cognitively undo the aversive event (Roese & Olson, 1995). Individuals who reported that they would not make any life choices differently experienced more overall life satisfaction than did individuals who would make different life choices in the work or family domains. These findings reflect the significance of life regrets in the aging years (Erikson, Erikson, & Kivnick, 1986). They also may reflect the valence of unfinished business (Savitsky, Medvec, & Gilovich, 1997), because the choice responses overwhelmingly indicated regrets over omissions rather than actions taken (see also Hattiangadi, Medvec, & Gilovich, 1995).

Congruent with traditional gender role norms, we found in Study 1 that men who reported that they had lived up to their intellectual abilities at midlife had particularly high occupational satisfaction 12 years later. Also, a slightly greater proportion of men than women said that they had lived up to their intellectual potential. In Study 2, more men than women who reported they had not lived up to their abilities responded that they would make work choices differently, and more men than women who reported they had lived up responded that there were no choices they would make differently. The pattern of responses in both studies reflects the vastly different opportunity structures confronting the men and women of the respondents' generation. Case materials in the Terman study suggest that these differences in opportunities were perceived by the Terman women (see Holahan, 1994; Holahan & Sears, 1995). Overall, the Terman men experienced considerable occupational success. Although the women's occupational achievements were superior to those of women of their cohort, they were modest in comparison with those of the Terman men (for more information on the Terman sample, see Holahan & Sears, 1995).

Some cautions should be noted in interpreting these results. Common method variance across measures (i.e., self-report questionnaires) may have contributed to the linkages between perceived regrets and life satisfaction (see Lecci, Okun, & Karoly, 1994). In addition, the findings of the present study are evidence of correlation only. Further, the Terman sample is uniquely advan-

taged, and attrition has made the sample somewhat more select in the areas of education and, for the men, occupational success (Holahan & Sears, 1995).

In summary, the present findings reflect the developmental significance of midlife self-appraisals (see Nelson, 1992; Levinson, 1978). They may also reflect the stability of adult personality, indicated by both trait approaches (e.g., Costa & McCrae, 1994) and studies of self-concept consistency (see Swann, 1997). However, although our results show a large amount of consistency over time, they should not be interpreted as suggesting that revision and life change after midlife are impossible. In fact, such revision can be accomplished either behaviorally or cognitively (see Nelson, 1992; Landman et al., 1995, for examples).

References

Campbell, A., Converse, P. E., & Rogers, W. I. (1976). *The quality of American life.* New York: Sage.

Costa, P. T., Jr., & McCrae, R. R. (1994). Set like plaster? Evidence for the stability of adult personality. In T. F. Heatherton & J. L. Weinberger (Eds.). *Can personality change?* (pp. 21-40). Washington, DC: American Psychological Association.

Erikson, E. H., Erikson, J. M., & Kivnick, H. Q. (1986). *Vital involvement in old age.* New York: Norton.

Hattiangadi, N., Medvec, V. H., & Gilovich, T. (1995). Failing to act: Regrets of Terman's geniuses. *International Journal of Aging and Human Development*, 40, 175-185.

Helson, R. (1992). Women's difficult times and the rewriting of the life story. *Psychology of Women Quarterly*, 16. 331-347.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94, 319-340.

Higgins, E. T., Bond, R. N., Klein, R., & Strauman, T. (1986). Self-discrepancies and emotional vulnerability: How magnitude, accessibility, and type of discrepancy influence affect. *Journal of Personality and Social Psychology*, 51. 5-15.

Holahan, C. K. (1984). Marital attitudes over forty years: A longitudinal and cohort analysis. *Journal of Gerontology*, 39, 49-57.

Holahan, C. K. (1988). Relation of life goals at age 70 to activity participation and health and psychological well-being among Terman's gifted men and women. *Psychology and Aging*. 3. 286-291.

Holahan, C. K. (1994). Women's goal orientations across the life cycle: Findings from the Terman Study of the Gifted. In B. Turner & L. Troll (Eds.). *Women growing older: Psychological perspectives* (pp. 35-67). Thousand Oaks, CA: Sage.

Holahan, C. K., & Sears. R. R. (1995). *The gifted group in later maturity.* Stanford. CA: Stanford University Press.

Idler, E. L., & Kasl, S. (1991). Health perceptions and survival: Do global evaluations of health status really predict mortality? *Journal of Gerontology*, 46, 55-65.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: User's guide.* Chicago: Scientific Software International.

Landman, J., & Manis, J. D. (1992). What might have been: Counterfactual thought concerning personal decisions. *British Journal of Psychology*. 83. 473-477.

Landman, J., Vandewater, E. A., Stewart, A. J., & Malley, J. E. (1995). Missed opportunities: Psychological ramifications of counterfactual thought in midlife women. *Journal of Adult Development*, 2. 87-97.

LaRue, A., Bank, L., Jarvik, L., & Hetland, M. (1979). Health in old age: How do physicians' ratings and self-ratings compare? *Journal of Gerontology*, 34, 687-691.

Lecci, L., Okun, M. A., & Karoly, P. (1994). Life regrets and current goals as predictors of psychological adjustment. *Journal of Personality and Social Psychology*, 66, 731-741.

Levinson, D. J. (with Darrow, C. N., Klein, E. B., Levinson, M. H., & McKee, B.). (1978). *The seasons of a man's life.* New York: Knopf.

Levinson, D. J. (with Levinson. J. D.). (1996). *The seasons of a woman's life.* New York: Knopf.

Markus, H., & Nurius. P. (1986). Possible selves. *American Psychologist,* 41, 954-969.

Markus, H., & Wurf. E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology.* 38. 299-337.

Martin, L. R., Friedman, H. S., Tucker, J. S., Schwartz. J. E., Criqui, M. H., Wingard, D. L.. & Tomlinson-Keasey. C. (1995). An archival prospective study of mental health and longevity. *Health Psychology*, 14, 381-387.

Miller, W. E., & Miller, A. H. (1977). *The CPS 1976 American National Election Study: Notes, frequencies, addendum and questionnaires.* Unpublished manuscript. Inter-University Consortium for Political and Social Research. Ann Arbor,MI.

Oden, M. H. (1968). The fulfillment of promise: 40-year follow-up of the Terman gifted group. *Genetic Psychology Monographs.* 77. 3-93.

Roese, N. J., & Olson, J. M. (Eds.), (1995). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 1-55). Mahwah, NJ: Erlbaum.

Ross, M., & Conway, M. (1986). Remembering one's own past: The construction of personal histories. In R. M. Sorrentino & E. T. Higgins (Eds.). *Handbook of motivation and cognition: Foundations of social behavior* (pp. 122-144). New York: Guilford Press.

Sauer, W. J., & Warland, R., (1982). Morale and life satisfaction. In D. J. Mangen & W. A. Peterson (Eds.), *Research instruments in social gerontology: Vol. 1. Clinical and social psychology* (pp. 195-210). Minneapolis: University of Minnesota Press.

Savitsky, K., Medvec, V. H., & Gilovich. T. (1997). Remembering and regretting: The Zeigarnik effect and the cognitive availability of regrettable actions and inactions. *Personality and Social Psychology Bulletin*, 23. 248-257.

Sears, P. S., & Barbee, A. H., (1977). Career and life satisfaction among Terman's gifted women. In J. C. Stanley, W. D. George, & C. H. Solano (Eds.). *The gifted and the creative: A fifty-year perspective* (pp. 28-65). Baltimore: Johns Hopkins University Press.

Sears, R. R, (1977). Sources of life satisfaction of the Terman gifted men. *American Psychologist*, 32, 119-128.

Swann, W. B. (1997). The trouble with change: Self-verification and allegiance to the self. *Psychological Science.* 8, 177-180.

Terman, L. M. (with Baldwin, B. T., Bronson, E., DeVoss, J. C., Fuller, F., Goodenough, F. L., Kelley, T. L., Lima, M., Marshall, H., Raubenheimer, A. H., Ruch, G. M., Willoughby, R. L., Wyam, J. B., & Yates. D. H.), (1925). *Mental and physical traits of a thousand gifted children: Vol. 1. Genetic studies of genius.* Stanford. CA: Stanford University Press.

Consequences of How We Define and Assess Intelligence

Wendy M. Williams, Cornell University

The author considers some of the consequences to society of how intelligence is defined and assessed. First, the author reviews historical approaches to understanding and measuring intelligence to clarify traditional points of view and current responses to these positions. Next, she describes a study of the Graduate Record Examination that illustrates the strengths and weaknesses of traditional approaches to assessing intelligence. She then describes a research program that defined, assessed, and trained intelligence from a different perspective — the perspective of practical intelligence. The article closes by considering directions for future research and thinking about a broader and more ecologically relevant conception of intelligence that would lead to new and potentially fruitful approaches to assessment and training.

What are the consequences of how our society defines and measures intelligence? Virtually everyone in the United States has been affected by prevailing views on the definition and assessment of intelligence. However, many people have never stopped to consider the impact of this issue on their lives. In this article, I discuss how the conceptualization of intelligence prevalent in the scientific community affects all of us, especially school children and college students. In some cases, these effects can be evaluated as being good versus bad; in other cases, they simply create advantages for certain groups of people possessing certain profiles of abilities. Any definition of intelligence carries with it a value judgment about the attributes and performances that are most prized by the society. In this discussion, I consider the logical consequences of the value judgments about intelligence that are emphasized by our society.

I begin by briefly reviewing background on historical approaches to understanding and measuring intelligence; my intention is to clarify the traditional points of view and current responses to these positions. Next, I describe a study that illustrates the strengths and weaknesses of traditional approaches to assessing intelligence, by portraying the consequences to graduate-school applicants of the use of the Graduate Record Examination (GRE). I follow by describing a research program that defined, assessed, and trained intelligence from a different perspective, and I discuss the implications of this approach. I close by considering directions for future research and thinking about the nature, definition, assessment, and training of intelligence.

Correspondence concerning this article should be addressed to Wendy M. Williams, Department of Human Development, Cornell University, Ithaca, New York 14853. Electronic mail may be sent via Intemet to wmw5@cornell.edu.

Reprinted from *Psychology and Aging*, Vol. 14, No. 2, 239-244, 1999.

Historical Approaches to Defining and Measuring Intelligence

A general definition of intelligence that most experts would accept is one that views intelligence as representing goal-directed, adaptive behavior. Two studies — one in 1921 and one in 1983 — asked experts to define intelligence (see Sternberg & Detterman, 1986). Common themes in the two groups of experts' opinions were the importance of learning from experience and the ability to adapt to the environment. In 1986, experts also mentioned the importance of people's understanding and control of their own thinking processes. Despite these apparent similarities among experts in their views of intelligence, early attempts to define and measure intelligence followed quite different trajectories.

The First Intelligence Tests

Two different traditions in the study of intelligence date back to the late 19th century to work by Sir Francis Gallon and Alfred Binet. Galton's (1883) psychophysical view of intelligence emphasized low-level tasks that tapped physical abilities in addition to mental abilities. Galton tested intelligence by measuring physical capabilities such as grip strength. However, scores on tests of grip strength and other physical abilities were not related to performance in school, a key domain in which people wanted to predict achievement (Wissler, 1901). Interest in Galton's views consequently waned. In some ways, however, Galton proved to be prescient: Later researchers, particularly Jensen (1982), discovered that when reliable measures (gathered using modern, reliable equipment) were developed to assess Galton's theory, they did correlate with the type of intelligence measured by Binet (discussed below) and others. However, following the early apparent disconfirmation by Wissler (1901), the Galtonian tradition remained largely unexplored for most of the 20th century; far more interest was shown in the perspective of Binet.

Binet approached the problem of defining intelligence very differently. He saw intelligence as consisting of direction (knowing what to do and how), adaptation (selecting a strategy for performing a task and monitoring one's success), and criticism (knowing how to critique one's work). In 1904, Binet and Theodosius Simon developed tests for the Paris school system that were designed to differentiate mentally defective children from children who were failing in school for other reasons. Binet's stated goal was to measure the abilities to judge well, comprehend well, and reason well (Binet & Simon, 1916). He developed the concept of mental age, which defined a child's intellectual performance compared with an average child of the same chronological age. Dividing mental age by chronological age (and multiplying by 100) results in a number called the IQ.

The next major improvement to the intelligence test occurred when Lewis Terman revised Binet and Simon's (1916) test, creating the Stanford-Binet

Intelligence Scales (Terman & Merrill, 1937), a set of tests that is still widely used today. The Stanford-Binet contains the following types of items: verbal reasoning, consisting of vocabulary, comprehension, absurdities, and verbal relations; quantitative reasoning, consisting of number series and arithmetic word problems; figural/abstract reasoning, consisting of pattern analysis; and short-term memory, consisting of memory for sentences, digits, and objects. Another major series of intelligence tests in common use today, developed by David Wechsler (1981), differs somewhat from the Stanford-Binet test. The Wechsler intelligence scales provide three scores: verbal, based on such subtests as Vocabulary and Verbal Similarities; performance, based on such subtests as Picture Completions and Picture Arrangements; and overall, which is a combination of the verbal and performance scores.

Binet's work a century ago defined a basic prototype for an intelligence test that remains substantially the same today. Scientists and others who use intelligence tests in research today are thus tacitly adopting a view of intelligence defined by Binet. Those who use modern versions of intelligence tests may not always realize that the use of these carries with it certain assumptions about the nature of intelligence, for example, that intelligence consists of the ability to respond quickly (e.g., some subtests award bonus points for speed) and solve mathematical problems. However, whenever a test is used to measure an ability, there is an assumption being made that the way the test constructor conceptualized that ability is reasonable. If a student takes a traditional intelligence test, scores well, and is consequently deemed very intelligent, we must bear in mind that intelligence in this sense means reading quickly and solving mathematical problems, for example. For the purposes of this dicussion, it is important to remember the types of assumptions that underlie the use of the widely used intelligence tests.

Psychometric Theories

All of this early interest in measuring intelligence was associated with the development of psychometric theories of intelligence, which view intelligence largely as a map of the mind (Sternberg, 1990). An early psychometric theory was proposed by Charles Spearman in 1904. Spearman used factor analysis to divide intelligence into what he called "*g*," or a single general factor, and multiple specific factors, each of which he called "*s*." How exactly is an estimate of *g* for a specific person obtained? The person completes a test such as the Wechsler or Stanford-Binet; next, the scorer conducts a principal-component analysis, which results in a first major factor representing internally consistent information about the test-taker's performance. This factor is called *g*.

g was seen by Spearman as a type of intelligence that influenced performance on all mental tests, whereas each *s* factor was thought to be involved in performance on a single type of test over and above the contribution made by *g*.

Spearman saw the general factor, *g*, as being at the heart of intelligence, and many researchers today would still agree with him (e.g., Gottfredson, 1986, 1996; Hunter & Schmidt, 1996). Subsequent theories concluded that the core of intelligence resided not in one factor but, rather, in multiple primary mental abilities (e.g., Cattell, 1971; Guilford, 1982; Thurstone, 1938; Vernon, 1971). These abilities included verbal comprehension, verbal fluency, inductive reasoning, spatial visualization, number, memory, and perceptual speed. In general, the key to all psychometric theories of intelligence is that they propose specific structures of intelligence explaining the organization of construct.

Information-Processing Theories of Intelligence

Information-processing theories focus on how people think and reason with their knowledge. Jensen (1982) looked at choice reaction time (how quickly a person can decide which button to push on a box); Hunt (1978) looked at lexical access speed (how fast people can retrieve information about words, or recognize the differences between pairs of letters such as AA, Aa, AB, and aB). Sternberg (1977a, 1977b) studied individual differences in intelligence by looking at how people solve verbal analogies. Simon (1976) looked at even more complex types of reasoning, such as those involved in playing chess. All of the information-processing approaches share an emphasis on the process of thinking and reasoning, rather than on the actual structure of intelligence.

Contemporary Systems Theories of Intelligence

Two contemporary theorists, Howard Gardner and Robert Sternberg, have proposed alternative ways to think about intelligence. Their views attempt to explain both how intelligence enables an individual to think and reason and how intelligence is structured. Gardner's (1983, 1993) theory of multiple intelligences proposed the existence of seven distinct intelligences, which can function alone or can interact to produce overall intelligent behavior — linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal. Gardner sees these seven intelligences as originating in different portions of the brain.

Sternberg's (1985, 1988) triarchic theory of intelligence emphasized a set of relatively interdependent processes. This theory postulates the existence of three important aspects of intelligence: componential, referring to information-processing components underlying intelligent performance (planning, monitoring, and evaluating performance; implementing one's plans; learning how to solve problems); experiential, which relates intelligence to experience; and contextual, which relates intelligence to everyday contexts (adaptation to, shaping of, and selection of environments). The systems theories of Gardner and Sternberg, as

well as those of other researchers, are, obviously, quite broad, and they have been criticized as being difficult to test fully or potentially to disconfirm, or both.

Implications for Assessment of the Different Theories of Intelligence

Suppose that a person accepts the psychometric view of intelligence and believes that *g*, or the general factor, is the best way to conceptualize meaningful intellectual ability. In such a case, measuring intelligence means measuring *g* (as just described). If a person believes that what is most relevant is a specific factor used in a specific type of performance, a test can be devised that measures success at this type of performance. However, if a person accepts a systems theory of intelligence, the task of assessing intelligence becomes quite different. In recognizing that intelligence is a complex process, systems theories necessarily define intelligence more broadly and make it more difficult to create a single test that would fairly measure intelligence as conceptualized within the systems view.

Most of the assessments, particularly the standardized assessments, used in North American schools are grounded in a relatively psychometric view of intelligence. In fact, as discussed above, the intelligence tests in wide use today are actually quite similar to the original tests developed by Binet and Simon in 1904. Tests such as the Scholastic Assessment Test (SAT) and Preliminary Scholastic Assessment Test (PSAT), the GRE, the Law School Admission Test (LSAT), and the Graduate Management Admission Test (GMAT), for example, all measure verbal and mathematical knowledge and reasoning. So far, there are no commercially available tests based on the systems view of intelligence that represent valid, reliable alternatives to the psychometric assessments currently in use. However, this does not mean that psychologists and educators should not evaluate the tests currently in use in order to advance the thinking about what constitutes intelligence and meaningful intellectual performance. In addition, by evaluating widely used tests and by determining where they may fall short, researchers may help to advance the development of better tests for future use.

How Good Are Current Tests at Predicting Real-World Performance?

I have discussed the origins of today's intelligence tests and the different theoretical perspectives that give rise to different types of intelligence tests. I have stated that intelligence tests were originally developed to predict school performance. But how good are the tests being used today? Do they predict meaningful aspects of intellectual performance in real-world environments? Are people who score high on intelligence tests more successful in general in their lives?

The degree to which intelligence tests predict out-of-school criteria such as job performance, for example, is a controversial question. Some believe that there is little or no justification for using tests of cognitive ability for job selection (McClelland, 1973). Others believe that cognitive ability tests are valid predictors of job performance for a wide variety of job settings (Barrett & Depinet, 1991) or even for all job settings (Schmidt & Hunter, 1981; see also Hawk, 1986; Gottfredson, 1986). Suppose that one accepts the pro-test position that stresses both the link between intelligence test scores and real-world performance and the fact that these scores are the best known predictors of job success. It is still the case that the majority of variance in real-world performance is not accounted for by intelligence test scores.

The average validity coefficient between cognitive-ability tests and measures of job performance is about .2 (Wigdor & Garner, 1982), meaning that test scores account for only four percent of the variance in job performance. The average validity coefficient between cognitive-ability tests and measures of performance in job training programs is about double (.4) that found for job performance itself, which suggests that the magnitude of prediction varies as a function of how comparable the criterion measure is to schooling. When the contexts are similar — as training is to sitting in a classroom — the prediction of cognitive ability is far greater.

Hunter and Schmidt have argued that validity coefficients should be corrected for unreliability in test scores and criterion measures and for restriction of range caused by the fact that only high scorers are hired. They believe that making these corrections results in better estimates of the true relation between cognitive-ability test performance and job performance, by raising the average validity coefficient to the level of about .5 (see, e.g., Hunter & Hunter, 1984; Schmidt & Hunter, 1981). This .5 value is hypothetical and is not routinely obtained in practice. But even the figure of .5 means that intelligence scores account for only 25 percent of the variance in job performance (i.e., the square of .5).

One might respond to the figure of 25 percent by thinking that traditional intelligence measures are not highly predictive and contain insufficient information to be of value in real-world decision making (e.g., in personnel hiring decisions). This is not necessarily true, however; Hunt (1995) has argued that substantial savings to employers can result from the use of currently available psychometric tests of intelligence to screen and select job applicants, even if the validity coefficients are very small. On the other side of this controversy is McClelland (1973), who has questioned the validity of cognitive-ability testing for predicting real-world criteria such as job performance. McClelland has argued in favor of competency tests that more closely reflect job performance itself.

Regardless of which side one wishes to endorse, it is clear that between 75 percent and 96 percent of the variance in real-world criteria such as job performance cannot be accounted for by individual differences in intelligence test scores. Thus, it would seem that scientists should be able to augment or improve upon the types of tests in use today, either by modifying current tests or by strengthening prediction by using additional tests that measure different types of aspects of intellectual performance.

One common cognitive-ability test, the GRE, is widely used in selecting applicants to matriculate in graduate school. The GRE is a well-known psychometric test that can be factored into a *g* factor and several *s* factors. Not surprisingly, the GRE is correlated moderately with IQ scores and all other *g*-saturated tests. The graduate-school environment is clearly a scholastic one, so presumably the GRE should be able to predict who will succeed in this environment. But, as anyone who has attended graduate school knows, there is more involved in succeeding in graduate school than just book-smarts. I now review a study I conducted with Robert Sternberg that evaluated the empirical validity of the GRE as a predictor of success in a graduate program in psychology (Sternberg & Williams, 1997).

Evaluating the GRE

Graduate programs use a variety of predictors to select those applicants who best match the programs and who offer the most to the field. An important element of every student's application is her or his score on the GRE. Students learn early that GREs are not to be taken lightly (literally or figuratively). Average scores are published in guides to graduate programs to help potential graduate students select "appropriate" programs before making an application. Some graduate programs list average scores of accepted students with the materials they send students to help them decide whether they should bother to apply at all. Other programs have taken the scores seriously enough to use them in a quantitative formula to help make admissions decisions (Dawes, 1971, 1975). Many programs have either explicit cutoffs or tacit minima, meaning that applicants who receive scores below these levels are almost never admitted.

Our study considered whether the GRE deserves its role, in light of its ability to predict who will succeed in graduate school. We evaluated the Verbal, Quantitative, and Analytical tests of the GRE, as well as the psychology subject-matter advanced achievement test. Our concern was with how the test is used, rather than with the test itself. The GRE has well-documented and adequate predictive validity with respect to certain criteria (see, e.g., Briel, O'Neill, & Scheuneman, 1993). The question we addressed was whether the criteria for which the GRE best predicts are the ones we care the most about. If the predicted criteria are secondary, then perhaps we need to seek additional forms of assessment.

How Is the GRE Used?

The admissions process is complex, and admissions committees must consider multiple factors. In the psychology department at Yale University, for example the site at which this research was conducted — GRE scores are just one of many factors considered. There are no explicit cutoffs, although the lower the level of the scores, the more an applicant needs compensating factors to gain admission. Applicants with very low scores are almost never admitted, regardless of compensating factors. At other institutions, GREs are used differently. In some fields at Cornell University, applications for admission are sorted upon arrival into four boxes, labeled "GRE Below 1200," "1200 to 1300," "1310 to 1400," and "Above 1400." This procedure is deemed necessary because some departments receive well over 100 applications for only eight admission places (for example), and the departments lack the personnel and time to read every application equally closely. In addition, the admissions committee has found over the years that there is very little variance in the strength and quality of letters of recommendation and undergraduate grade point average (GPA), leaving the students' personal statements and GREs as the main sources of variation.

In past research on testing, Robert Sternberg and I described a factor called the publication reason, which increases reliance on tests such as the GRE (e.g., Sternberg, 1988; Williams et al., 1996). When average test scores such as GREs (and SATs, LSATs, GMATs, etc.) are published, there is pressure upon university personnel to keep these average scores high to remain competitive with other institutions in the public eye. However, if a department admits only students with high GREs, department personnel may come to believe that high GREs are essential for success in their graduate program, as everyone who succeeds has high GREs. If no students with low or moderate GREs are ever admitted, it becomes impossible to falsify the view that high GREs are essential for success.

Another point admissions committees must consider is that many of the sources of financial support for entering students are university wide. The students who receive this financial aid are often selected solely on the basis of GREs, because there are no other criteria that can be directly and fairly compared across academic fields (i.e., all of the students nominated for such fellowships possess high GPAs and glowing letters of recommendation). It can, thus, be difficult for departments to secure funding for students with low GREs. Hence, GRE scores find their way into the selection process from several different angles.

Why Study the GRE?

In general, the GRE is used extensively in admissions and financial-aid decision making, alongside other factors and types of information. It is also used by government agencies, such as the National Science Foundation, as one factor in awarding graduate fellowships. For students, the GRE and preparation for it are

expensive, time consuming, and potentially anxiety provoking. The GRE yields scores that are taken as indications of intellectual abilities, and students usually take these scores seriously. In addition — and important for the purposes of our discussion — the GRE may mispredict performance in a graduate program, unfairly disadvantaging some students and advantaging others.

The GRE is based on conventional psychometric notions of abilities (discussed above), which traditionally have tended to emphasize some abilities (e.g., verbal, quantitative, and analytical), arguably at the expense of other abilities (e.g., creative, practical; see Sternberg, 1985, 1988, 1996). The GRE is also an example of psychological theory put into practice and raises practical problems of prediction or its lack thereof. Finally, if GRE scores are not sufficiently valid for the kinds of decisions for which they are used, mental contamination may result in their being used anyway, as long as they are available (T. D. Wilson & Brekke, 1994). In other words, knowing a student's GRE scores may tacitly influence admissions decisions, even if the scores are acknowledged to be of limited value.

What Has Past Research Revealed About the Usefulness of the GRE?

Dunlap (1979) studied admissions criteria as predictors of academic performance and professional potential of social-work students. Performance was successfully predicted from the faculty interview and undergraduate GPA. The GRE was a weak predictor, and letters of reference were of little value. Other studies have shown that GREs can be good predictors of grades and faculty evaluations, at least for first-year graduate performance in psychology (Dawes, 1971). Clearly, the test has been demonstrated to have some predictive validity to some criteria of success in graduate school; however, an informal review of 149 studies on the predictive validity of the GRE across fields showed that, on average, the GRE considered alone accounted for a little less than 10 percent of the variation in the various criteria of graduate performance (Wood & Wong, 1992).

According to the GRE Technical Manual (Briel et al., 1993), although a three-factor solution fits the verbal, quantitative, and analytical portions fairly well, the items tend to be rather highly correlated. Rock, Werts, and Grandy (1982) found that the verbal and quantitative factors were correlated .64, the verbal and analytical items .77, and the quantitative and analytical items .77. A slightly better fit of model to data was obtained when a reading-comprehension item was separated from the verbal item (see also Powers & Swinton, 1981). Similar correlations between pairs of items have been found by others (e.g., Schaeffer & Kingston, 1988).

Empirical validities of the GRE vary somewhat by field. K. M. Wilson (1979) showed that, in the prediction of first-year grades, the median validity coefficients for first-year grades in psychology graduate school were .18 for the

Verbal test, .19 for the Quantitative test, and .32 for the subject-matter test. The tendency for the subject-matter test to predict first-year grades better than the Verbal, Quantitative, and Analytical tests is common. In another cooperative study, the median correlation of the subject-matter test with first-year grades was .31 (Burton & Turner, 1983). In this same study, predictive validities of the GRE to first-year grades for the Verbal and Quantitative tests over all social sciences were .26 and .22, respectively. Schneider and Briel (1990) found overall correlations with first-year grades of .26 for the Verbal test, .25 for the Quantitative test, .24 for the Analytic test, and .36 for the subject-matter test. Undergraduate GPA showed a similar correlation to that for the subject-matter test (.34). Recent research focusing on graduate training in physics revealed that GREs provide only marginal prediction of graduate performance. In addition, there are sex differences in scores on the Physics Advanced Test (in favor of men) that are not reflected in graduate-school performance (Glanz, 1996).

Some studies have focused on criteria other than first-year grades in graduate school. Rock (1974) found that for applicants for National Science Foundation fellowships, correlations of GRE scores with attainment of the PhD in psychology in two random samples were .12 and .19 for the Verbal test, .33 and .14 for the Quantitative test, and .19 and .24 for the subject-matter test. Schrader (1978, 1980) obtained citation counts from the Social Sciences Citation Index and from the Annual Reviews of Psychology as well as publication rates from the Psychological Abstracts. He found correlations with these criteria of .15 to .30 for the GRE Verbal, .24 to .32 for the GRE Quantitative, and .32 to .47 with the GRE subject-matter advanced test. But in another study, Clark and Centra (1982) failed to detect any significant correlations between publication rates and GRE scores for recent PhDs.

Some studies have focused on relative predictions for various subpopulations of graduate-school students. Braun and Jones (1985) found no differential prediction across subgroups varying in age, sex, or race; however, Swinton (1987) found significant underprediction of first-year grade averages for women in all fields of graduate study.

In related research focusing on the SAT, Crouse and Trusheim (1991) argued that the selection benefits colleges derive from using the SAT in admissions decisions are minimal (see, also, Crouse & Trusheim, 1988; Jencks & Crouse, 1982). Similarly, in research focusing on the Medical College Adminissions Test (MCAT), Gough and Hall (1975) studied the prediction of academic versus clinical performance in medical school. They found that academic performance was predicted by the MCAT and premedical GPA. However, clinical performance was not predicted from MCAT scores and premedical academic-achievement indices. Gough and Hall noted that the clinical performance factor was more important than academic attainment in explaining who excelled in medical school.

Why Don't These Tests Better Predict Who Will Succeed in College and Graduate School?

The fact that the tests used to select applicants do not robustly predict who will succeed in college or graduate school raises the issue of academic versus practical or real-world problems (e.g., Neisser, 1976; Neisser et al., 1996). Neisser (1976) was one of the first psychologists to press the distinction between academic and practical intelligence. He described academic-intelligence tasks (common in the classroom and on intelligence tests) as formulated by others, often of little or no intrinsic interest, having all needed information available from the beginning and disembedded from an individual's ordinary experience. In addition, these tasks usually are well defined, have but one correct answer, and often have just one method of obtaining the correct solution (Wagner & Sternberg, 1985). Note that these characteristics apply less well to many of the problems we face in our daily lives, especially at work. Work problems often are unformulated or in need of reformulation, of personal interest, lacking in information necessary for solution, related to everyday experience, poorly defined, characterized by multiple "correct" solutions (each with liabilities as well as assets), and characterized by multiple methods for picking a problem solution.

The distinction between academic intelligence ("book smarts") and practical intelligence ("street smarts") has long been recognized by the nonscientist. Many common expressions attest to the essential role of practical intelligence in everyday life (e.g., "learning the ropes" and "getting your feet wet"). Both laypeople and researchers include concepts of academic and practical intelligence in their own implicit theories of intelligence (Sternberg et al., 1981). Recently, practical intelligence has been the focus of a growing number of studies carried out in a wide range of settings and cultures. Summaries of aspects of this literature have been provided by Ceci (1996), Rogoff and Lave (1984), Scribner and Cole (1981), Sternberg and Wagner (1986, 1994), Sternberg, Wagner, and Okagaki (1993), and Voss, Perkins and Segal (1991).

The distinction between academic and practical intelligence is illustrated by studies in which participants were assessed on both academic and practical tasks. The consistent result is little or no correlation between performance on the two kinds of tasks. IQ is unrelated to the order-filling performance of milk-processing plant workers (Scribner, 1986); the degree to which racetrack handicappers use a complex and effective algorithm (Ceci & Liken 1986, 1988); the complexity of strategies used in computer-simulated roles such as city manager (Dorner & Kreuzig, 1983; Dorner, Kreuzig, Reither, & Staudel, 1983); and the tacit knowledge of undergraduates (Wagner, 1987; Wagner & Sternberg, 1985), business managers (Wagner & Sternberg, 1990), salesperons (Wagner, Rashotte, & Sternberg, 1992), and Air Force recruits (Eddy, 1988). In addition, the accuracy with which grocery shoppers identified quantities of food that pro-

vided the best value per price was unrelated to their performance on the M.I.T. mental arithmetic test (Lave, Murtaugh, & de la Roche, 1984; Murtaugh, 1985).

The distinction between academic and practical problems can help us understand why the GRE and tests like it fall short in predicting performance in real-world environments (as I will discuss below). It is not surprising that standardized tests tend better to predict later performance on academic tasks than they do on practical or real-world tasks. In the case of the medical school research (Gough & Hall, 1975), the MCAT predicted later performance of an academic type, but not the essential clinical performance of medical-school students, which involves solving different types of problems from those found on standardized tests and course exams. For these and other related reasons, researchers have been critical of the entire psychometric approach (e.g., Ceci, 1996; Gardner, 1983, 1993; Neisser et al., 1996).

Partly in response to the psychometric tradition, a growing body of research has focused on delineating and assessing the types of ability needed to succeed on practical as opposed to academic problems (see, e.g., Wagner, 1987, Wagner & Sternberg, 1985, Williams & Sternberg, 1993; Williams et al., 1996; for a review, see Sternberg, Wagner, Williams, & Horvath, 1995). Much of this research on practical problems is related to the systems-theory view of intelligence, discussed earlier. The psychometric view of intelligence, and specifically the view that *g* is the single best measure of intelligence, is often seen as being at odds with current systems views. Members of the psychometric camp tend to point out that the GRE and tests like it provide the best estimates available of the type of ability needed for graduate school; for example, if *g* is what matters to success, then people scoring high on the GRE should do better in graduate school than people who do not score high, given the high *g* loading of the GRE.

Systems theorists, on the other hand, would argue that the abilities measured by the GRE are only a subset of the abilities needed in graduate school. According to this view, the GRE is seen as being of limited value in predicting who will succeed in the graduate-school environment. For these reasons, systems theorists such as Ceci (1996) have called for empirical verification of the predictive validity of the GRE and other similar *g*-loaded tests and have expressed concern that even if these tests do predict worthwhile outcomes, it is essential to reveal the mechanisms through which they do so (i.e., to distinguish between description and explanation), as well as the contextual factors relevant to these predictions.

After reviewing the literature on what selection tests such as the GRE do and don't do, one can conclude that these tests usually provide a modest amount of information about first-year course performance in graduate or professional school. But this fact raises the question of whether we should be selecting students on the basis of who will get the best grades in first-year courses. Ultimately, what really matters is not first-year grades, but meaningful performance as a psychologist, in graduate school and thereafter. How well does the

GRE predict performance as a professional psychologist, when administered prior to starting a graduate program? These were the questions we sought to answer by studying the performance of graduate students in the Yale University psychology department.

What Did We Predict, and Why?

On the basis of Sternberg's (1985) triarchic theory of intelligence (discussed above), which views intelligence in terms of three types of abilities (analytic, creative, and practical), we predicted that analytical abilities would be the most important for performance on the GRE, in view of its emphasis on factual recall, and would predominate in course performance as well, given the way courses tend to be taught. Thus, we expected GREs to provide some prediction of course grades. The theory also suggests that practical and especially creative abilities will be critical for performance as a psychological researcher or practitioner and that although analytical abilities will also be important, they will not hold any privileged position. Moreover, even these abilities will be within a domain (psychology) as practiced within a particular context (e.g., university, private practice), so that the validity of the GRE for a broad array of domains would be open to question.

Introductory graduate psychology courses usually require the same kinds of fairly abstract and often context-lean memorization and analysis that are required by conventional tests of ability and achievement. Thus, the GRE, as a measure of analytical ability, would be likely to predict course-grades. However, success in psychology as a career, or even in the latter years of graduate training in psychology, may require creative and practical abilities in addition to analytical abilities. Creative abilities are necessary to formulate theories, empirical research, or hypotheses. Practical abilities are required to succeed within the university promotion system; to get grants funded; or to attract, keep, and treat clients. Thus, we might expect the GRE to be weaker as a predictor of success as a psychologist, or even as an advanced graduate student, than we would expect it to be as a predictor of initial graduate grades. This weaker relation follows from the fact that the GRE measures primarily analytical abilities. Such abilities, especially as measured in the context of a standardized test, will be only minimally related to creative and practical abilities (see Sternberg, 1985, 1996; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996).

We also predicted that (a) the best predictor of graduate grades would be the GRE advanced (subject-matter) test in psychology, because the best predictor of future achievement of a given kind is past achievement of the same kind; (b) the GRE would be a better predictor of first-year graduate grades than of second-year grades, because the testing is closer to the first year of performance and because first-year courses are less advanced (the more advanced the courses become, the more they are likely to draw on skills beyond the conventional

memory and analytical ones); and (c) the GRE would be only weakly predictive or not predictive of more meaningful criteria of graduate-program success, such as ratings by professors of various aspects of the quality of students and their work, including ratings of students' demonstrated analytical, creative, practical, research, and teaching abilities, as well as of their dissertations. If the GRE predicted anything, we expected it to be best for professors' ratings of analytical ability. However, we believed that the correlation would not necessarily be substantial, because analytical ability as demonstrated in the actual graduate work context might be somewhat different from analytical ability as demonstrated in a paper-and-pencil test.

Method

To test our hypotheses, we asked all faculty members in the Department of Psychology at Yale University who supervised graduate students or who were on dissertation committees during the period from 1980 to 1991 (N = 40) to provide certain ratings. Graduate advisors were asked to rate their primary graduate-student advisees (i.e., those for whom they were Ph.D. dissertation supervisor) for five types of abilities: analytical, creative, practical, research, and teaching. Evaluations were on a low (1) to high (7) scale for each rating. The faculty members were told to use the scale in the following way: 7 = absolutely superlative — among the very best in our graduate program; 6 = outstanding — among the top 10 percent in our graduate program; 5 = excellent — among the top 25 percent in our graduate program; 4 = very good-among the top 50 percent in our graduate program; 3 = good — among the top 75 percent in our graduate program; 2 = fair — among the top 90 percent in our graduate program; and 1 = poor — among the very weakest in our graduate program.

Obviously, there is some degree of subjectivity in ratings such as these, and the possibility of halo effects as well.

To broaden our criterion information, we obtained the overall evaluations of dissertations (for those students who had completed dissertations) from the three ratings given by the three dissertation-committee readers (all of whom were psychology faculty but none of whom was the primary dissertation advisor). These ratings were on a 4-point scale. Because lower ratings corresponded to higher evaluations of dissertations, we reflected the ratings for this study to make the direction consistent with all the other measures (higher numbers corresponding to better performance). We also computed GPAs for students' first year, second year, and combined first and second years of graduate training. These GPAs were based on a grading system of honors (4), high pass (3), pass (2), and fail (0), with failing grades being exceedingly rare. We included in our sample all matriculants, including those who had not completed the program.

We used as predictor variables scores from the Verbal, Quantitative, Analytical, and advanced tests of the GRE. The verbal, quantitative, and analyti-

cal sections are required at Yale; the advanced test is not. Our sample ultimately consisted of 170 graduate students. Of these students, 84 had completed their dissertations at the time of our study, meaning that for these 84 students we were able to obtain the dissertation readers' evaluations of their dissertations. Data on 3 students were incomplete, leading to 167 students in the final sample (68 men, 99 women). Because the GRE advanced test is not required at Yale, and because some students take the test in a field other than psychology (resulting in the exclusion of their advanced test scores from the sample), the number of advanced test scores was reduced ($N = 73$).

Results

There was a good range in GRE scores among the students in the psychology department: Verbal test scores ranged from 250 to 800 ($M = 653$, $SD = 97$), Quantitative test scores from 320 to 840 ($M = 672$, $SD = 78$), Analytical test scores from 410 to 810 ($M = 656$, $SD = 92$), and psychology advanced scores from 490 to 850 ($M = 690$, $SD = 65$). Advisor ratings ranged from 1 to 7; mean advisor ratings ranged from 4.46 on creative ability to 4.75 on analytical ability, and standard deviations ranged from 1.45 to 1.65. Year 1 GPA ranged from 2.46 to 4.00 ($M = 3.58$, $SD = .31$), and Year 2 GPA ranged from 2.90 to 4.00 ($M = 3.72$, $SD = .33$). Dissertation reader ratings (the mean of three readers per student) ranged from 1 to 3.33 ($M = 3.10$, $SD = .56$). Although the mean GRE scores in the mid to high 600s were relatively high and well above the national average, the Verbal, Quantitative, and Analytical test standard deviations and ranges were also quite high. The standard deviations came close to the national ones. Ratings of student performance also varied considerably. Thus, restrictions of range cannot be blamed for our results, a point I will return to later.

When we looked at the results separately for men and women, we found no significant differences on any of the measures: Men and women were comparable in measured abilities and performance in our sample. We found that the separate GRE scores were related to each other, with intercorrelations ranging from .17 to .58 (for men and women combined). We also looked at intercorrelations within the various criteria (averaging the two years of GPA). Overall GPA during the first two years correlated modestly with the other criteria: .41 with analytical rating ($p < .001$), .16 with creativity rating ($p < .05$), .29 with practical rating ($p < .001$), .29 with research rating ($p < .001$), and .32 with teaching rating ($p < .001$), as well as .32 with mean dissertation reader rating ($p < .01$).

The most important analyses were the correlations of the predictors (GRE scores) with the criteria (advisors' and readers' ratings as well as grades). These data are shown in Table 1, for the sexes combined and separately for men and women. First, GREs did have some modest value for predicting grades, at least in the first year of graduate study. The median correlation across the four scores for men and women conbined was .17.

Table 1
Correlations of GRE Scores with Criteria

GRE subtest	Year 1 GPA	Year 2 GPA	Overall GPA	Analytical rating	Creative rating	Practical rating	Research rating	Teaching rating	Dissertation reader rating
GRE Verbal									
Overall	.18* (159)	.10 (152)	.17* (159)	.12 (161)	.14 (161)	.12 (161)	.12 (161)	.15 (157)	.08 (83)
Men	.16 (63)	.04 (61)	.12 (63)	.11 (66)	.21 (66)	.09 (66)	.15 (66)	.20 (63)	.07 (34)
Women	.21* (96)	.14 (91)	.21* (96)	.13 (95)	.09 (95)	.15 (95)	.10 (95)	.10 (94)	.08 (49)
GRE Quantitative									
Overall	.14 (159)	-.01 (152)	.09 (159)	.16* (161)	-.04 (161)	-.02 (161)	.07 (161)	-.04 (157)	.07 (83)
Men	.20 (63)	-.05 (61)	.10 (63)	.26* (66)	.16 (66)	-.02 (66)	.18 (66)	.07 (63)	-.10 (34)
Women	.08 (96)	.03 (91)	.07 (96)	.06 (95)	-.21 (95)	-.02 (95)	-.04 (95)	-.13 (94)	.19 (49)
GRE Analytical									
Overall	.17* (155)	.03 (149)	.12 (155)	.18* (157)	.16* (157)	.05 (157)	.12 (157)	.14 (154)	.24* (80)
Men	.24 (61)	.06 (59)	.21 (61)	.31** (64)	.36** (64)	.04 (64)	.30** (64)	.28* (62)	.47** (32)
Women	.11 (94)	.00 (90)	.06 (94)	.05 (93)	-.01 (93)	.05 (93)	-.03 (93)	.01 (92)	.05 (48)
GRE Psychology									
Overall	.37*** (71)	.02 (70)	.23 (71)	.18 (70)	.12 (70)	.07 (70)	.14 (70)	.01 (66)	.15 (45)
Men	.40* (32)	.02 (31)	.24 (32)	.16 (32)	.11(32)	.10 (32)	.21 (32)	.05 (29)	.12 (22)
Women	.37* (39)	.01 (39)	.22 (39)	.21 (38)	.15 (38)	.05 (38)	.09 (38)	.01 (37)	.17 (23)

Note. Numbers in parentheses are sample sizes. GRE = Graduate Record Examination; GPA = grade point average. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Second, we had assumed a better prediction of first-year grades than of second-year grades in graduate study. In fact, although three of the four correlations for men and women combined were statistically significant for Year 1, none of the correlations for Year 2 was statistically significant (with a median across the four correlations of just .02). The psychology advanced test showed a significantly higher correlation with Year 1 GPA than with Year 2 GPA (with
significantly higher correlation with Year 1 GPA than with Year 2 GPA (with
listwise deletion for missing cases), $t(67) = 2.48$, $p = .02$ (see Cohen & Cohen, 1983, p. 56, for the formula used to test for the significance of the difference between these two dependent correlations). However, the differences between the correlations for Year 1 and Year 2 GPA for the other GRE subtests did not reach statistical significance.

Third, the psychology advanced test correlated strongly with Year 1 GPA (.37). Test-measured achievement was thus a strong predictor of a grade-based achievement (as we predicted). Consequently, the GREs did provide modest prediction of grades in the first year; however, grades represent one of the least important aspects of graduate performance. We did not find prediction to the second year of grades.

Fourth, with one exception, the GREs were not useful as predictors of other aspects of graduate performance: ratings of analytic, creative, practical, research, and teaching abilities by primary advisors and ratings of dissertation quality by faculty readers. For the combined sexes, only 4 of 24 correlations reached statistical significance. The median was only .12. We had expected that correlations would be higher with the analytical than with the advisors' other ratings, but given the level of the correlations, there just was not enough relation for many of the correlations to be significantly different from zero, much less from each other. Thus, as systems views of intelligence would predict, GREs were generally not valid or otherwise useful predictors of important aspects of success.

Fifth, it turns out that the four statistically significant correlations for the combined sexes cannot quite be taken at face value, because the table shows that in every case the effect is due to correlations for men, but not for women. There was, in fact, one consistently successful (although only marginally significant) predictor of ratings: the GRE Analytical test score for men only ($z = 1.95$, $p < .06$, two-tailed; see Cohen & Cohen, 1983, p. 53, for the formula used to test for the significance of the difference between the independent correlations for men and women). (The only other significant correlations were for the GRE Quantitative test score predicting advisor's analytical rating for men, and for the GRE Quantitative test score predicting advisor's creative rating for women, but in the negative direction.)

We also investigated several other data-analytic approaches that had the potential to improve the GRE's predictive power. For example, we conducted multiple regressions that optimally combined the various GRE scores, both with

and without the advanced test. The results were substantially unchanged: Combination of multiple scores did not significantly improve prediction. We also did canonical regressions, thereby using multivariate analysis linearly to combine dependent as well as independent variables in an optimal way. Again, we obtained no significant improvements in prediction.

In sum, GREs were found to be modest predictors of first-year but not second-year grades in our graduate program, both for men and for women. However, only the GRE Anaytical test score was found to predict more consequential evaluations of student performance, and only for men.

Criticisms of Our Study

One potential criticism of our study concerns the issue of range restriction. However, restriction of range cannot be fully blamed for the pattern of correlations we obtained. First, as noted earlier, our standard deviations and ranges were rather substantial. Second, the fact that significant correlations were obtained between GREs and first-year grades, and between the GRE Analytical test score for men and the professors' ratings, suggests that correlations could be obtained where they existed. Third, good prediction of grades was found for the GRE subtest with the lowest scaled-score standard deviation (the psychology advanced test).[1]

A second criticism concerns the unreliability of faculty ratings. Some might argue that any kind of subjective rating is notoriously unreliable, so that one could hardly expect any test, the GRE included, to show substantial correlations with such unreliable and possibly invalid criteria. In fact, grades correlated with the ratings, the ratings correlated with each other, and the GRE Analytical test score correlated with the ratings for men. Consequently, it was possible to obtain correlations with the ratings, suggesting unreliability was not responsible for the failure of the GRE to correlate with the ratings.

A third criticism is that Yale University graduate students are not typical of graduate students in general. It is possible that our findings might not replicate for other programs and other graduate students. It is also possible that Yale's graduate-program emphasis might also be unrepresentative. However, like most graduate programs, Yale's clinical training emphasizes the Boulder model of the scientist-practitioner, and Yale's nonclinical track trains students for industry and government positions as well as for traditional academic jobs. Furthermore, the range in GRE scores of admitted applicants provides additional testimony to

---

[1]Although we tried to correct for restriction of range, we acknowledged that Yale's students are not typical of all students who enter psychology graduate programs and that different results might be obtained across the entire range of students in all programs. We also acknowledged that our relatively small sample size restricted both the power of the significance tests to detect actual differences or relationships and the generalizability of our results to future psychology graduate students at Yale and at other schools, as well.

the diversity of the student body. (Note that lack of proficiency in English by foreign students was rare in Yale's psychology department; but, nevertheless, the analysis performed on GRE Quantitative test scores alone addresses the possibility that our results would change if we ignored verbal proficiency in English.)

Conclusions of Our Study

The results of our study underscored the need for serious validation studies of the GRE, not to mention other admissions indices, against measures of consequential performances, whether of students or of professionals. The point is that we should apply the same standards of falsifiability in our admissions process as we do in our scientific work. Sometimes, the use of a test can become self-perpetuating, without serious attempts to verify its effectiveness. Our study suggested the need to reflect on our use of tests before they become firmly and even irrevocably entrenched. Psychologists regularly create, refine, and export standardized tests for use throughout the academic community and our society in general. Thus, psychologists should remain aware of the issues revealed by this study and work to ensure that the tests we advocate are used effectively and appropriately.

An obvious direction for the tests of the future is to expand upon the types of abilities or intelligences assessed (e.g., practical and creative intelligence). Theories such as Sternberg's (1985) triarchic theory, Ceci's (1996; Ceci & Roazzi, 1994) bio-ecological theory, or Gardner's (1983, 1993) multiple intelligences theory might be used as bases for the development of expanded tests, both paper-and-pencil and performance-based. To guide this future work, psychologists need to remain reflective about what is meant by the concept of intelligence as it applies in the context of success in a graduate program in psychology (see Neisser, 1979; Sternberg, 1985).

In conclusion, this research found the traditional psychometric approach to predicting graduate-school achievement, represented by GRE, to be somewhat lacking in prediction of meaningful aspects of success in graduate school. Like other studies that have examined related *g*-loaded intelligence tests (the GRE, SAT, PSAT, LSAT, and GMAT), we found that the GRE is best at predicting grades earned in the semesters immediately following admission. Thus, in general, when we select students for admission and financial aid awards based on high GRE scores, we are selecting students most likely to do well in course work but not necessarily more likely to do well in research and teaching than applicants with lower GRE scores. This, in a sentence, is the likely implication for graduate students of the psychometric view of intelligence and the use of tests constructed on the basis of this view. Further research is warranted so that we can develop better predictive tools that provide more and better information about meaningful aspects of performance. Perhaps some of these tests of the

future will be based on systems-oriented views of intelligence. I now discuss one example of an educational program that used a systems approach as the basis for instruction and assessment.

One Educational Program Based on a Broader Conception of Intelligence

What happens when one applies a systems theory of intelligence to classroom instruction and assessment? To investigate this question, Howard Gardner, Tina Blythe, Noel White, and Jin Li at Harvard University collaborated with Robert Sternberg and myself at Yale University on the Practical Intelligence for School (PIFS) Project (Williams et al., 1996, 1997). This work began with the observation that possession of analytical or academic intelligence does not always lead to success in school. Children also need practical and creative thinking skills. In the world after graduation from school, practical and creative skills are likely to become even more essential to success (see Sternberg et al., 1995).

Consider an experience shared by many parents: A child with good reading and writing skills and a solid vocabulary hands in a messy composition, filled with cross outs, after insisting to her parents that her teacher said that what really counts are the child's ideas. A week goes by, and the child receives a poor grade — and is shocked. This child is intelligent in the traditional, *g*-based sense and does well on tests of intelligence (see Neisser, 1979; Neisser et al., 1996). Yet, the child seems to lack some kind of intelligence relevant to the school environment, what my associates and I called "practical intelligence for school" (Gardner, Krechevsky, Sternberg, & Okagaki, 1994; Sternberg, Okagaki, & Jackson, 1990; Williams et al., 1996). Students with practical intelligence for school both understand and are able to respond appropriately to the demands of the school environment, which include doing homework, taking tests, reading for understanding, and writing effectively.

Practical intelligence for school is a specific aspect of the more general construct of practical intelligence, which has been studied by a variety of investigators in a range of contexts (see Ceci, 1996; Sternberg & Wagner, 1986; Sternberg et al., 1995, for reviews). This research has shown that people have a set of procedural-knowledge skills that are relevant to their adaptation to real-world environments. This set of procedural-knowledge skills are now well or fully conceptualized by conventional notions of intelligence and are not well measured by conventional intelligence tests. Some do not accept that practical intelligence exists (e.g., Ree & Earles, 1993; Schmidt & Hunter, 1993). However, my associates and I believed that there was sufficient evidence for such a construct that it was worth pursuing its applications to the classroom environment. Practical intelligence can be viewed as a part of intelligence, broadly defined, whether or not it is viewed as wholly distinct from the academ-

ic aspects of intelligence.

Can Intelligence Be Taught?

Given the fuzzy nature of this question, it is not surprising that the scientific evidence has been mixed and subject to alternative interpretations. On the one hand, a number of controlled studies have yielded impressive gains. For example, the ODYSSEY project was designed to raise the intellectual skills and school performance of Venezuelan school children of roughly middle-school age and was evaluated with highly favorable results by Herrnstein, Nickerson, DeSanchez, and Swets (1986). Ramey (1994; Ramey & Campbell, 1984, 1992; Ramey et al., 1992), studying younger children, also accumulated substantial evidence that gains in intelligence and school performance are possible as a result of intensive interventions among high-risk preschoolers. Other programs have also shown at least limited, and sometimes quite impressive, success (Bereiter & Engelmann, 1966; Feuerstein, 1980; Garber, 1988; Nisbett, Fong, Lehman, & Cheng, 1987; Schweinhart, Barnes, & Weikart, 1993; see, also, Detterman & Sternberg, 1982; Honig, 1994; Nickerson, 1994; Nickerson, Perkins, & Smith, 1985).

Some believe, however, that it is not possible to increase intelligence. Herrnstein and Murray's (1994) review of the literature led them to conclude that little meaningful gain is possible; others have come to the same conclusion (Jensen, 1969, 1989; McLaughlin, 1977; Spitz, 1986, 1992). In the middle of the road are studies that are encouraging, but cautious, in their interpretations (Consortium for Longitudinal Studies, 1983; Lazar & Darlington, 1982; Snow & Yalow, 1982; Zigler & Berman, 1983). In sum, no serious psychologist has suggested that unlimited gains in intelligence are possible; however, some believe that modest to moderate gains are possible in some cases under limited circumstances.

Theoretical Motivation

The theoretical motivation for our research was a combination of the theory of multiple intelligences proposed by Gardner (1983, 1993) and the triarchic theory of intelligence proposed by Sternberg (1985, 1996). We integrated the two theoretical frameworks by viewing intelligence in the seven domains proposed by Gardner as having the three aspects proposed by Sternberg. Consider, for example, the case of linguistic intelligence, an important component of school success. Linguistic intelligence may be seen as encompassing analytical aspects (e.g., in understanding how to develop a logically consistent argument), creative aspects (e.g., in writing a creative essay or a poem), and practical aspects (e.g., in writing or speaking persuasively to one's teacher or fellow students). The same merging applies to each of the multiple intelligences.

In this set of studies, we sought to boost school achievement by creating an intervention that would develop practical intelligence for school for middle-school-age students (Williams et al., 1996). The PIFS program is a curricular intervention designed to enhance the practical-thinking skills of fifth- and sixth-grade students. We focused on fifth- and sixth-grade students because we believed that the point at which a child leaves primary school and enters intermediate school is a time when the child is ripe for instruction in practical thinking skills. The child at this juncture is old enough to assimilate and use the skills, but young enough to be open to learning them. Also, the child's practical thinking skills will become more essential as she or he enters the intermediate school environment, in which she or he must change classes several times a day and deal with the demands of different teachers. The PIFS Project involved intensive observations and interviews of students and teachers to determine the tacit knowledge necessary for success in school.

How did we uncover the practical thinking skills essential to school success? We began by developing a taxonomy of practical thinking skills. This taxonomy consisted of five themes — knowing why, knowing self, knowing differences, knowing process, and reworking — which were applied to practical thinking in four domains — reading, writing, homework, and test taking. To illustrate, consider the taxonomy applied to skills in test taking: The first theme is knowing why. Students master this area of the test-taking curriculum by answering questions such as What are the roles of tests in and out of school? How does testing relate to other class work? The second theme is knowing self, for which students learn how to recognize current study strategies and test-taking practices and identify personal strengths and weaknesses in terms of testing.

The third theme is knowing differences, for which students learn how to recognize different kinds of tests and test questions, within and across subjects, learn what each test can and cannot determine about the test taker, and learn different strategies that are appropriate for each test. For the fourth theme, knowing process, students come to understand that long-term preparation is necessary to preparing for tests, and they learn both long-term and short-term strategies for test preparation, as well as strategies for solving problems during actual test taking. The fifth and final theme, reworking, involves students' using the results of tests as an opportunity for self-reflection and as a stepping stone toward more productive learning and test taking.

The PIFS curriculum book, distributed to teachers, contained an overview of the scientific basis of the research, teacher training materials, and 35 one- to three-hour-long lessons that could be adapted by teachers to their students' needs. The curriculum was implemented twice over two consecutive years in schools in Connecticut (n = 193 students) and Massachusetts (n = 321 students) in a matched-control-group design. We developed pre- and posttests designed to assess the quality of students' practical knowledge in each of four focal areas

covered by the curriculum (reading, writing, homework, and testing). These tests were administered in October and June. All of the tests were based on the kinds of tasks students are typically asked to do in school to make them fair to students who had not been exposed to the curriculum.

For example, we used two 50-minute reading assessments, one based on a factual passage and the other on a passage of fiction. Students read the passage and answered questions about their general understanding, their thinking processes while reading, the parts they found easy or hard to understand and why, how they would study for a test on the passage, and so on. Most of the questions were open-ended. The writing assessment involved two parts, each of which lasted 50 minutes. The first part of the pretest asked students to write a composition describing in detail a place they knew well. For the posttest, students described a person they knew well.

Following the writing, students answered questions about their writing process — what was easy or hard, how they got their ideas and organized their presentation, and what their teacher's reaction might be. For the second part of the writing assessment, students revised the composition they wrote on the first part. They then reflected on the revision process, indicating the parts they had added or deleted, explaining those changes, and predicting what the teacher might and might not like about the piece. Each assessment thus yielded traditional academic measures of intelligence (e.g., the grammatical correctness of an essay) along with measures of practical intelligence (e.g., the persuasiveness of an essay).

Results

The result of PIFS curriculum evaluations were uniformly positive. Analyses of covariance (ANCOVAs) were conducted on all academic- and practical-intelligence variables, using the pretest score for each measure as the covariate, and comparing the Fall-to-Spring score changes for the PIFS and control-group students. In general, the PIFS program successfully enhanced practical and academic skills. Positive results were observed in both years of the program in Studies 1 and 2 at Connecticut and Massachusetts sites. The PIFS effect occurred across a variety of initial conditions in which the PIFS group began either lower than, equal to, or higher than the control group.

Consider representative results from the Year 2 Connecticut-site data (these results are quite similar to the Massachusetts-site results). The means for individual variables (academic and practical measures of reading, writing, homework, and test-taking ability) ranged from 2.29 (SD = .70) to 4.09 (SD = .69) on the 5-point rating scale, ranging from poor (1) to excellent (5). For academic writing ability, for example, the PIFS (treatment) group showed a significant pretest-to-posttest increase, $t(51) = 6.67, p < .001$. The control group also

increased, but the change was not significant, $t(49) = 1.08$. An ANCOVA showed that pretest score was a significant covariate and that there was a significant PIFS effect, covariate $F(1, 99) = 60.22$, $p < .001$. The PIFS group's gains were significantly greater than the control group's gains, $F(1, 99) = 16.49$, $p < .001$.[2] For practical writing ability, for example, both the PIFS and control groups showed significant increases from pretest to posttest, respectively, $t(51) = 9.89$, $p < .001$, and $t(49) = 2.61$, $p < .05$. An ANCOVA showed that pretest score was a significant covariate and that there was a significant PIFS effect, covariate $F(1, 99) = 88.19$, $p < .001$. The PIFS group's gains were significantly greater than the control group's gains, $F(1, 99) = 25.33$, $p < .001$.[3] For an overall practical-intelligence measure, the combined summary score showed both PIFS and control groups with significant pretest-posttest gains, respectively, $t(53) = 14.32$, $p < .001$, and $t(51) = 3.24$, $p < .01$. An ANCOVA showed that the pretest score was a significant covariate and that there was a significant PIFS effect, covariate $F(1, 103) = 214.61$, $p < .001$. Once again, the PIFS group's gains were significantly greater than the control group's gains, $F(1, 103) = 60.89$, $p < .001$[4]

Thus, in general, we found that the PIFS program successfully enhanced both practical and academic skills in each of the target skill areas (reading, writing, homework, and test taking) in children from diverse socioeconomic backgrounds attending diverse types of schools. In addition, teachers, students, and administrators alike reported fewer behavioral problems in PIFS classes.

Conclusions

This research demonstrated that the role of tacit knowledge in school success is central, and significantly, this research showed that tacit knowledge can be effectively defined, efficiently taught, and used by students to improve their per-

[2]The mean covariate-adjusted improvement from the mean pretest score (of the whole sample) was .66 points for PIFS participants and .12 points for the control participants. This represented a 26 percent increase at posttest for the PIFS participants compared with a 5 percent increase for control participants. After removing variance in posttest scores accounted for by pretest scores, the PIFS treatment variable accounted for 10 percent of the total variance in posttest scores and 14 percent of the variance in regressed change.

[3]The mean covariate-adjusted improvement from the mean pretest score (of the whole sample) was .82 points for PIFS participants and .24 points for the control participants. This represented a 32 percent increase at posttest for the PIFS participants compared with a 9 percent increase for control participants. After removing variance in posttest scores accounted for by pretest scores, the PIFS treatment variable accounted for 12 percent of the total variance in posttest scores and 20 percent of the variance in regressed change.

[4]The mean covariate-adjusted improvement from the mean pretest score (of the whole sample) was .76 points for PIFS participants and .20 points for the control participants. This represented a 29 percent increase at posttest for the PIFS subjects compared with an 8 percent increase for control participants. After removing variance in posttest scores accounted for by pretest scores, the PIFS treatment variable accounted for 17 percent of the total variance in posttest scores and 37 percent of the variance in regressed change.

formance in school. This is welcome news for students who fall short of their potential because they lack basic practical insights into their teachers' expectations and how to fulfill these expectations. For teachers, the possibility of training practical intelligence for school may mean less frustration with students who do not perform because of an array of factors not related to lack of analytical ability. The PIFS curriculum provides one method for and approach to training these essential practical skills. Students exposed to PIFS become better able to make optimal use of their gifts and abilities within the context of the school environment, while learning practical skills they can use throughout their lives.

Over the past several decades there have been various trends favoring different types of curricular approaches and interventions. Sometimes, curricula are developed and implemented on the basis of anecdata, or anecdotal reports, instead of carefully controlled studies. For example, a recent trend in educational intervention has focused on building emotional and moral intelligence (e.g., Coles, 1996). Lately, it is often said that we must educate our youth for character and moral values (e.g., the Character Counts curriculum in use across the United States). Certainly, we are not suggesting that there is anything wrong with wanting children to develop character and morality. However, the questions of exactly how to accomplish this goal (or any other educational goal), and of how to know if we are succeeding with any individual program, can only be answered in a scientifically adequate way through empirical research.

Unlike the latest fad that sometimes becomes a curricular intervention in the absence of a solid theoretical foundation and rigorous supporting data, the PIFS program is rooted in theory and based on hard empirical evidence. Our data show something meaningful and promising by providing evidence that it is possible to improve broad-based intellectual skills. By focusing on reading, writing, homework, and test-taking ability, we cast our net wide in an attempt to create meaningful changes in broad areas of students' intellectual performances.

As I have already discussed, there has been widespread disagreement regarding the degree to which children's intellectual capabilities can be modified. For example, Herrnstein and Murray (1994) essentially dismissed intervention effects, arguing that short of adoption, there are no meaningful ways to raise the intellectual performance of children. Although future research is needed to assess the long-term durability of training in practical intelligence, our data showed reasonable increases over the school year. Thus, on the topic of the controversy regarding potential intervention effects, we weigh in on the side of cautious optimism. In the very least, our data suggest that further research on increasing practical intelligence is warranted.

A broader issue raised by this research concerns the definition of intelligence itself and how one's definition affects one's viewpoint regarding the modifiability of intelligence and how best to enhance it. On the theoretical side, a growing literature suggests that traditional *g*-based psychometric conceptions of intelligence are incomplete. Interest in the type of intelligence people use to solve

real-world problems, referred to here as practical intelligence, has led to a broad cross section of studies identifying practical intelligence in different domains (Ceci, 1996; Rogoff & Lave, 1984; Scribner & Cole, 1981; Sternberg & Wagner, 1986, 1994; Sternberg et al., 1993; Voss et al., 1991) and even studies showing that practical intelligence can be assessed and taught (see Sternberg et al., 1995, for a review).

Thus, theoretically speaking, a psychometrically based measure of intelligence such as *g* may not be the whole story when it comes to understanding children's intelligence. Practically speaking, psychometrically based *g* may not be sufficient if we wish to be fair and accurate in our assessment of children's capabilities in the classroom and beyond. Just as Renzulli (1986) has argued for a broad-based approach to identifying giftedness — in which children are identified on the basis of not only above-average ability, but also high levels of motivation and creativity — we argued for a broad-based approach to identifying school-based competence. We believed that practical intelligence should be seen as an essential component of children's competence, worthy of assessment and instruction in its own right. Our research suggests that training in practical intelligence can help children remediate areas of weakness, as well as build on existing skills, to improve their performance in many academic areas.

The evidence regarding the modifiability of *g*-based intelligence is mixed. But if we can accept that intelligence is more than *g*, there is hope that meaningful increases in intelligence can be achieved, even if, for example, these increases do not focus on *g*-based abilities. Thus, putting aside one's point of view regarding whether we can affect measures of *g* through training, our research shows that we can affect measures of practical intelligence through training. Whether one wishes to call practical intelligence a type of intelligence, a type of knowledge, a set of skills, or whatever, the point is that it can be delineated and taught successfully.

The training of practical intelligence may be particularly useful in challenged populations, because students in these populations may have had little opportunity to acquire school-relevant practical intelligence on their own or at home. Training challenged students may help them to overcome a deficit in school-related knowledge and skills that could otherwise have derailed them. Many of these students may have latent capabilities that they have not harnessed or profited from because of a lack of fit between the student and the school environment. By helping students understand what is expected of them and why, and by demystifying the process of succeeding in school, training in practical intelligence may help to reach students who have previously opted out of the school experience.

In conclusion, this research showed that practical intelligence can be identified, assessed, and taught, in order to achieve meaningful increases in real-world success in the classroom. What are the implications of the broader definition of intelligence that provided the theoretical motivation of this research? First, our

broad definition of intelligence led us to develop instructional approaches focusing on broader competencies in the reading, writing, homework, and test-taking domains than the competencies typically focused on in traditional instruction. In particular, we focused on helping students to understand the reasons for the work they were asked to do and to develop insights into their strengths and weaknesses, as well as into the processes involved in doing good work. Our assessments mirrored these competencies and delved deeper and more explicitly into the nature of students' thinking process than do most traditional *g*-based assessments. It is important that the students who received our training did better not only on measures of practical intelligence but also on measures of academic intelligence. Thus, my associates and I are optimistic about the potential of a broad conception of intelligence — one based on meaningful intellectual accomplishments — to enrich education and improve assessment. We advocate further research on practical intelligence for school to broaden our appreciation and understanding of this construct and its applications in the classroom.[5]

Where Should We Be Headed in Future Work on Intelligence?

I began this article by noting that the intelligence tests in use today are based on psychometric conceptions of intelligence that stress analytical reasoning ability in verbal, mathematical, and figural domains, and that can be distilled into a single measure of intellectual functioning, *g*. An enormous literature and reams of data testify to the ability of these tests to predict different types of performances. Few would argue with the fact that traditional measures of intelligence are valuable tools to measure certain types of abilities relevant in certain environments. However, as was discussed, even the best estimates offered of the predictive power of traditional intelligence tests claim that these tests account for 25 percent of the variance in performance. Clearly, a substantial amount of unexplained variance exists, and to the extent that this variance is systematic, it is possible that a broader perspective on intelligence may enable us to assess these previously unmeasured aspects of intelligence.

There are two assumptions made by most researchers working in the psychometric tradition. These assumptions are, first, that intelligence can be conceptualized and measured as a single general factor that predicts success in a wide variety of tasks and environments and, second, that intelligence is not trainable

[5]We have also adapted ideas from the PIFS Project to train students' creative skills (Sternberg & Williams, 1996; Williams, Brigockas, & Sternberg, 1997). Once again, the idea motivating our attempt to train creative skills is a belief in the importance of creative skills in intellectual behavior. In this research, we developed a curriculum containing lessons focusing on helping students meet the following types of goals: building self-efficacy, questioning assumptions, defining and redefining problems, generating many ideas and cross fertilizing these ideas across participants, engaging in sensible risk-taking, tolerating ambiguity, accepting mistakes, surmounting obstacles, delaying gratification, and collaborating with other creative people.

(see, e.g., Herrnstein & Murray, 1994). In this article, I presented data questioning both of these assumptions. The GRE study showed that the general factor, *g*, was not predictive of success in graduate school in this case study. The PIFS Project showed that there are aspects of intelligence in addition to *g* that are important to success in school and that these aspects of intelligence can be assessed and taught. It thus appears that there is sufficient evidence to proceed with research, with the goals of expanding our conceptualization of intelligence and honing our ability to train broad-based types of intelligence.

The PIFS Project was one attempt to base instruction and assessment on a broader conception of intelligence. The results of this project provided grounds for optimism about the potential for broader conceptions of intelligence and assessments that throw their nets more widely than do traditional psychometric *g*-based measures of intelligence. Moving beyond the school environment, it is possible that success in real-world managerial job environments may be predicted by measures of practical intelligence or that success in artistic environments may be predicted by measures of creative thinking abilities. How might these alternative types of intelligence be measured?

In recent research, I measured practical intelligence and examined its ability to predict success in business management (Williams & Sternberg, 1997). Practical intelligence was assessed through the use of scenarios depicting managerial problems for which the test taker generated or rated potential responses.[6] Sternberg and Lubart (1995) measured creative intelligence by asking individuals to write short stories and artistic compositions with novel themes, to create advertisements for dull products such as doorknobs and bow ties, and to solve scientific problems such as how we might be able to tell whether someone has been on the moon within the past month. To measure multiple intelligences, Gardner (1993) advocated the use of portfolio and performance assessments. Say, for example, that the goal is to improve upon the prediction provided by the GRE. Some undergraduates already submit samples of their work when they apply to graduate school, and it would not be a major step to require a somewhat more formal portfolio of work, which could then be evaluated in terms of criteria believed to be most relevant to graduate success (with appropriate safeguards against plagiarism.)

Certainly, I am not advocating broad use of alternative types of assessments before these assessments have been carefully evaluated in multiple applications in a variety of environments. Sensibly used, alternative types of intelligence tests are still more likely to supplement than to replace conventional tests.

---

[6]These responses were scored by comparing them with expert-generated profiles of ideal responses. By correlating practical-intelligence measures with external and independent criteria of success, I showed that practical intelligence was associated with success. For example, practical intelligence was related to the success measures compensation ($r = .39$, $p < ,001$), age-controlled compensation ($r = .38$, $p < .001$), and level of position ($r = .36$, $p < .001$), even after controlling for background and educational experience.

However, the scientific community should remain open-minded about the possibility that alternative types of tests may help us to improve upon the prediction of performance based solely on traditional tests. In conclusion, I am not suggesting that the use of intelligence tests be abandoned; what I am suggesting is that we attempt to augment the traditional forms of these tests with assessments that measure a broader and more ecologically relevant set of abilities and evaluate the increment to prediction of real-world success provided by such an expanded test battery.

References

Barren, G.V., & Depinet, R. L. (1991). A reconsideration for testing for competence rather than for intelligence. *American Psychologist, 46,* 1012-1024.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool.* Englewood Cliffs, NJ: Prentice-Hall.

Binet, A., & Simon, T. (1916). *The development of intelligence in children.* Baltimore: Williams & Wilkins.

Braun, H. I., & Jones, D. H. (1985). *Use of Empirical Bayes method in the study of the validity of academic predictors of graduate school performance* (GRE Board Professional Rep. GREB No. 79-13P, ETS Rep. 84-34). Princeton, NJ: Educational Testing Service.

Briel, J. B., O'Neill, K., & Scheuneman, J. D., (Eds.). (1993). *GRE technical manual.* Princeton, NJ: Educational Testing Service.

Burton, N., & Turner, N. (1983). *Effectiveness of Graduate Record Examinations for predicting first year grades: 1981-82 summary report of the Graduate Record Examinations Validity Study Service.* Princeton, NJ: Educational Testing Service.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development* (expanded ed.). Cambridge, MA: Harvard University Press.

Ceci, S. J., & Liken J. (1986). Academic and nonacademic intelligence: An experimental separation. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 119-142). New York: Cambridge University Press.

Ceci, S. J., & Liken J. (1988). Stalking the IQ-expertise relationship: When the critics go fishing. *Journal of Experimental Psychology: General. 117,* 96-100.

Ceci, S. J., & Roazzi, A. (1994). The effects of context on cognition: Postcards from Brazil. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 74-101). New York: Cambridge University Press.

Clark, M. J., & Centra, J. A. (1982). *Conditions influencing the career accomplishments of Ph.D.'s* (GRE Board Research Rep. GREG No. 76-2R; ETS Research Rep. 82-18). Princeton, NJ: Educational Testing Service.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coles, R. (1996). *The moral education of children.* Cambridge, MA: Harvard University Press.

Consortium for Longitudinal Studies. (1983). *As the twig is bent: Lasting effects of preschool programs.* Hillsdale, NJ: Erlbaum.

Crouse, J., & Trusheim, D. (1988). The case against the SAT. *Public Interest, 93,* 97-110.

Crouse, J., & Trusheim, D. (1991). How colleges can correctly determine selection benefits from the SAT. *Harvard Educational Review, 61,* 125-147.

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26,* 180-188.

Dawes, R. M. (1975). Graduate admission variables and future success. *Science. 187,* 721-723.

Detterman, D. K., & Sternberg, R. J. (Eds.). (1982). *How and how much can intelligence be increased?* Norwood, NJ: Ablex.

Dorner, D., & Kreuzig, H. (1983). Problemlosefahigkeit and intelligenz [Problem-solving and intelligence]. *Psychologische Rundschaus, 34,* 185-192.

Dorner, D., Kreuzig, H., Reither, F., & Staudel, T. (1983). *Lohhausen: Yom Umgang mit Unbestimmtheir and*

Kornplexitat. Bern, Switzerland: Huber.

Dunlap, W. R. (1979). How effective are graduate social work admission criteria? *Journal of Education for Social Work, 15, 96-102.*

Eddy, A. S. (1988). *The relationship between the Tacit Knowledge Inventory for Managers and the Armed Services Vocational Aptitude Battery.* Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Feuerstein, R. (1980). *Instrumental enrichment.* Baltimore: University Park Press.

Galton, F. (1883). *Inquiry into human faculty and its development.* London: Macmillan.

Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk.* Washington, DC: American Association on Mental Retardation.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gardner, H., Krechevsky, M., Sternberg, R. J., & Okagaki, L. (1994). Intelligence in context: Enhancing students' practical intelligence for school. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 105-127). Cambridge, MA: MIT Press.

Glanz, J. (1996). How not to pick a physicist. *Science, 274, 710-712.*

Gottfredson, L. S. (1986). Societal consequences of the *g* factor in employment. *Journal of Vocational Psychology, 29, 379-410.*

Gottfredson, L. S. (1996). Racially gerrymandering the content of police tests to satisfy the U.S. Justice Department: A case study. *Psychology, Public Policy, and Law, 2, 418-446.*

Gough, H. G., & Hall, W. B. (1975). The prediction of academic and clinical performance in medical school. *Research in Higher Education, 3, 301-314.*

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89, 48-59.*

Hawk, J. (1986). Real world implications of g. *Journal of Vocational Behavior, 29, 411-414.* .

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Herrnstein, R. J., Nickerson, R. S., DeSanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist, 41, 1279-1289.*

Honig, A. S. (1994). Intervention, infant and preschool. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence (Vol. 1, pp. 599-607). New* York: Macmillan.

Hunt, E. B. (1978). Mechanics of verbal ability. *Psychological Review, 85, 109-130.*

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce.* New York: Russell Sage Foundation.

Hunter, J. E., & Hunter, R. F (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-98.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2,* 447-472.

Jencks, C., & Crouse, J. (1982). Aptitude vs. achievement: Should we replace the SAT? *Public Interest, 67,* 21-35.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1-123.

Jensen, A. R. (1982). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 255-310). Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1989). Raising IQ without increasing g? A review of "The Milwaukee Project: Preventing mental retardation in children at risk." *Developmental Review, 9,* 234-258.

Lave, J., Murtaugh, M., & de la Roche, O. (1984). The dialectic of arithmetic in grocery shopping. In B. Rogoff & J. Lace (Eds.), *Everyday cognition: Its development in social context* (pp. 67-94). Cambridge, MA: Harvard University Press.

Lazar, I., & Darlington, R. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (2-3, Serial No. 195).

McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist, 28,* 1-14.

McLaughlin, D. H. (1977). *Title I, 1965-1975: A synthesis of the findings of federal studies.* Palo Alto, CA: American Institutes for Research.

Murtaugh, M. (1985, Fall). The practice of arithmetic by American grocery shoppers. *Anthropology and Education Quarterly, 16,* 186-192.
Neisser, U. (1976). General, academic, and artificial intelligence. In L. Resnick (Ed.), *Human intelligence: Perspectives on its theory and measurement* (pp. 179-189). Norwood, NJ: Ablex.
Neisser, U. (1979). The concept of intelligence. *Intelligence, 3,* 217-227.
Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, W. A., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77-101.
Nickerson, R. S. (1994). Interventions, later. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 1, pp. 607-613). New York: Macmillan.
Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking.* Hillsdale, NJ: Erlbaum.
Nisbett, R. F., Fong, G. T., Lehman, D. R., & Cheng, P W. (1987). Teaching reasoning. *Science, 238,* 625-631.
Powers, D. E., & Swinton, S. S. (1981). Extending the measurement of graduate admission abilities beyond the verbal and quantitative domains. *Applied Psychological Measurement, 5,* 141-158.
Ramey, C. T. (1994). Abecedarian project. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. I, pp. 1-3). New York: Macmillan.
Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88,* 515-523.
Ramey, C. T., & Campbell, F. A. (1992). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A. Huston (Ed.), *Children in poverty* (pp. 190-221). New York: Cambridge University Press.
Ramey, C. T., Campbell, E. A., Bryant, D. M., Burchinal, M., Sparling, J., & Wasik, B. H.(1992). Early intervention and long-term predictors of school status. In F. Parker et al. (Eds.), *New directions in child and family research* (pp. 170-174). Washington, DC: Administration on Children, Youth, and Families.
Ree, M. J., & Earles, J. A. (1993). *g* is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Sciences. 1,* 11-12.
Renzulli, J. S. (1986). The three ring conception of giftedness: A developmental model for creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 53-92). New York: Cambridge University Press.
Rock, D. A. (1974). *The prediction of doctorate attainment in psychology, mathematics, and chemistry* (GRE Board Research Rep. 69-6aR). Princeton, NJ: Educational Testing Service.
Rock, D. A., Werts, C., & Grandy, J. (1982). *Construct validity of the GRE Aptitude Test across populations — An empirical confirmatory study* (GRE Board Professional Rep.No. 78-IP; also ETS Research Rep. 81-57). Princeton, NJ: Educational Testing Service.
Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: Its development in social context.* Cambridge, MA: Harvard University Press.
Schaeffer, G. A., & Kingston, N. M. (1988). *Strength of the analytical factor of the GRE in several subgroups: A full-information factor analysis approach* (GRE Board Professional Rep. No. 86-7P; also ETS Research Rep. 88-5). Princeton, NJ: Educational Testing Service.
Schmidt, F. L., & Hunter,. J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36,* 1128-1137.
Schmidt, F. L., & Hunter, J. E. (1993). Tactic knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 1,* 8-9.
Schneider, L. M., & Briel, J. B. (1990). *Validity of the GRE: 1988-1989 summary report.* Princeton, NJ: Educational Testing Service.
Schrader, W. B. (1978). *Admissions test scores as predictors of career achievement in psychology* (GRE Board Research Rep. No. 76-1R). Princeton, NJ: Educational Testing Service.
Schrader, W. B. (1980). *GRE scores as predictors of career achievement in history* (GRE Board Research Rep. No. 76-1R). Princeton, NJ: Educational Testing Service.
Schweinhart, L. J., Barnes, H. V., & Weikart, D. P (1993). *Significant benefits: The High/Scope Perry Preschool Study through age 27* (Monograph No. 10). Ypsilanti, MI: High/Scope Press.
Scribner, S. (1986). Thinking in action: Some characteristics of practical thought. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 13-30). New York: Cambridge University Press.
Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.
Simon, H. A. (1976). Identifying basic abilities underlying intelligent performance of complex tasks. In L. B.

Simon, H. A. (1976). Identifying basic abilities underlying intelligent performance of complex tasks. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 65-98). Hillsdale, NJ: Erlbaum.

Snow, R. E., & Yalow, R. (1982). Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493-585). New York: Cambridge University Press.

Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence.* Hillsdale, NJ: Erlbaum.

Spitz, H. H. (1992). Does the Carolina Abecedarian early intervention project prevent sociocultural mental retardation? *Intelligence, 16,* 225-237.

Sternberg, R. J. (1977a). Component processes in analogical reasoning. *Psychological Review, 84,* 353-378.

Sternberg, R. J. (1977b). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence.* New York: Viking.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1994). Diversifying instruction and assessment. *Educational Forum,* 59(1), 47-53.

Sternberg, R. J. (1996). *Successful intelligence.* New York: Simon & Schuster.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41,* 37-55.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly, 40,* 129-137.

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity.* New York: Free Press.

Sternberg, R. J., Okagaki, L., & Jackson, A. (1990). Practical intelligence for success in school. *Educational Leadership, 48,* 35-39.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world.* New York: Cambridge University Press.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1994). *Mind in context.* New York: Cambridge University Press.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacit knowledge in work and at school. In H. Reese & J. Puckett (Eds.), *Advances in lifespan development* (pp. 205-227). Hillsdale, NJ: Erlbaum.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50,* 912-927.

Sternberg, R. J., & Williams, W. M. (1996). *Developing creativity in students and teachers.* Washington, DC: Association for Supervision and Curriculum Development.

Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training *of* psychologists? A case study. *American Psychologist, 52,* 630-641.

Swinton, S. S. (1987). *The predictive validity of the restructured GRE with particular attention to the older students* (GRE Board Professional Rep. No. 83-25P; also ETS Research Rep. 87-22). Princeton, NJ: Educational Testing Service.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* Boston: Houghton Mifflin.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University *of* Chicago Press.

Vernon, P E. (1971). *The structure of human abilities.* London: Methuen.

Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.). (1991). *Informal reasoning and education.* Hillsdale, NJ: Erlbaum.

Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology, 52,* 1236-1247.

Wagner, R. K., Rashotte, C. A., & Sternberg, R. J. (1992). *Tacit knowledge in sales: Rules of thumb for selling anything to anyone.* Unpublished manuscript.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits. The role of tacit knowledge. *Journal of Personality and Social Psychology, 49,* 436-458.

Wagner, R. K., & Sternberg, R. J. (1990). Street smarts. In K. E. Clark & M. B. Clark (Eds.), *Measures of lead-*

ership (pp. 493-504). West Orange, NJ: Leadership Library of America.
Wigdor, A. K., & Garner, W. R. (Eds.), (1982). *Ability testing: Uses, consequences, and controversies.* Washington, DC: National Academy Press.
Williams, W. M., Blythe, T., White, N., Li, J., Gardner, H., & Sternberg, R. J. (1997).*Practical intelligence for school: Evaluation of a program for improving the academic potentials of middle-school students.* Manuscript submitted for publication.
Williams, W. M., Blythe, T., White, N., Li, J., Sternberg, R. J., & Gardner, H. (1996). *Practical intelligence for school.* New York: HarperCollins.
Williams, W. M., Brigockas, M., & Sternberg, R. J. (1997). *Creative intelligence for school.* New York: Addison Wesley.
Williams, W. M., & Sternberg, R. J. (1993). Seven lessons for helping children make the most of their abilities. *Educational Psychology, 13,* 317-331.
Williams, W. M., & Sternberg, R. J. (1997). *Success acts for managers.* San Diego, CA: Harcourt Brace Professional.
Wilson, K. M. (1979). *The validation of GRE scores as predictors of first-year performance in graduate study: Report of the GRE cooperative validity studies project* (GRE Board Research Rep. No. 75-85R). Princeton, NJ: Educational Testing Service.
Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.
Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Review Monograph* (6, Suppl. 3).
Wood, J. L. & Wong, A. C. (1992). GRE scores and graduate school success. *International News and Notes. p. 1.*
Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised.* San Antonio, TX: The Psychological Corporation.
Zigler, E., & Berman, W. (1983). Discerning the future of early childhood intervention. *American Psychologist, 38,* 894-906.

The Locus of Adult Intelligence: Knowledge, Abilities, and Nonability Traits

Phillip L. Ackerman and Eric L. Rolfhus, Georgia Institute of Technology

Some intelligence theorists (e.g., R. B. Cattell, 1943; D. O. Hebb, 1942) have suggested that knowledge is one aspect of human intelligence that is well preserved or increases during adult development. Very little is known about knowledge structures across different domains or about how individual differences in knowledge relate to other traits. Twenty academic and technology-oriented tests were administered to 135 middle-aged adults. In comparison with younger college students, the middle-aged adults knew more about nearly all of the various knowledge domains. Knowledge was partly predicted by general intelligence, by crystallized abilities, and by personality, interest, and self-concept. Implications of this work are discussed in the context of a developmental theory that focuses on the acquisition and maintenance of intelligence-as-knowledge, as well as the role of knowledge for predicting the vocational and avocational task performance of adults.

In the history of modern psychology, assessment of intelligence started as an enterprise solely devoted to children and adolescents (Binet & Simon, 1908/1916). Modern assessment of adult intelligence has changed few of the fundamental properties of the Binet-Simon approach. The Binet-Simon paradigm involves "(a) Tests of higher-order mental processes; (b) Elimination (as far as possible) from consideration of knowledge acquired via specific instruction; (c) Elicitation of maximal effort on the part of the examinee [in contrast to typical effort]; and (d) School achievement as the fundamental criterion for external validation" (Ackerman & Heggestad, 1997, p. 221). Whether this is most appropriate for adult assessment, both conceptually and from an applied standpoint, has not been answered.

Adult intelligence research after World War I took two specific perspectives on aging. The first approach was to simply describe changes in adult levels of performance on various intelligence scales (e.g., Gilbert, 1935). The general

---

This research was supported by a grant from The College Board and by Grant F49620-98 from the Air Force Office of Scientific Research. Portions of this work were presented in College Board Research Report 98-3 (see Ackerman & Rolfhus, 1998). We thank Wayne Camara for his encouragement and assistance, on behalf of The College Board, in supporting this research — especially in making available the College Level Examination Program and the Advanced Placement tests. We also gratefully acknowledge the assistance of the Kanfer-Ackerman Laboratory staff for assisting in developing test items, running the experiments, and coding the data.

Correspondence concerning this article should be addressed to Phillip L. Ackerman, School of Psychology, Georgia Institute of Technology, Psychology Building-0170, 274 5th Street, Atlanta, GA 30332-0170. Email may be sent to phillip.ackerman@psych.gatech.edu.

Reprinted from *Psychology and Aging*, Vol. 14, No. 2, 314-330, 1999.

sense from these studies was that there is a long and substantial decline of intelligence during adulthood, starting with adults in their mid-20s. This work was contrasted with the results of longitudinal studies indicating that IQs of individual participants generally were well-preserved across adulthood (e.g., Bayley, 1955) — thus pointing to cohort differences as the explanation for the age-related declines in intelligence.

In the second approach to aging and intelligence, other researchers focused on redefining the issue by changing the focus of what intelligence actually means for middle-aged and older adults (e.g., Miles, 1934; see also Demming & Pressey, 1957, for an early test of practical intelligence). In these and other studies, researchers identified aspects of the traditional IQ test that were not relevant to adult intelligence, or they attempted to place increased emphasis on verbal abilities and decreased emphasis on general reasoning abilities.

Later efforts by Schaie and Strother (1968), using Thurstone's Primary Mental Abilities (PMA) test battery, found that the overwhelming trend (except for speeded tests) was for the longitudinal data to support increasing intelligence as age increased, up to age 60 or so, whereas the cross-sectional data showed that the younger participants outperformed the older participants. Several follow-up data waves have been collected, culminating in Schaie's (1996) recent review of the project. Schaie reported that "Verbal Meaning, Space, and Reasoning attain a peak plateau in midlife from the 40s to the early 60s, whereas Number and Word Fluency peak earlier and show very modest decline beginning in the 50s" (p. 351). This work represents a most comprehensive review — but it remains firmly grounded in an approach to intelligence that is not fundamentally different from that of Binet and Simon.

Theory

Researchers have used two similar theoretical perspectives to attempt to explain and understand the nature of adult intellect — the perspectives of Hebb (1941, 1942) and of Cattell (1941, 1943; see Ackerman, 1996, for a review). Hebb differentiated two types of intelligence, A and B, where Intelligence A (physiologically based) peaked in early adulthood and then declined as age increased, and where Intelligence B (educational and experiential) was well-preserved through much of adult life. Intelligence A was most sensitive to neurological incidents and disease, whereas Intelligence B was most robust to such effects. Cattell similarly differentiated fluid intelligence (Gf) from crystallized intelligence (Gc), where Gf was physiologically based, and Gc was educationally and experientially based. Cattell linked Gf/Gc theory into a wide nomological network, resulting in an "investment hypothesis" that included interests, personality, and motivational traits (see Cattell, 1971). Note that the term trait used throughout this article refers to a transituational disposition on which individuals may differ (Matthews & Deary, 1998). The term trait corresponds to Snow's (1989) usage of the term aptitude.

There has been extensive discussion in the literature over the past three decades about the plausibility of Cattell's (1943) Gf/Gc distinction (e.g., Baltes & Schaie, 1976; Horn, 1989; Horn & Cattell, 1966). Nonetheless, there is some general consensus that Gf abilities are more likely to show age-related declines in adulthood, whereas Gc abilities are less likely to show age-related declines in adulthood (at least up to age 60 or 70). The major shortcoming to the whole discussion of Gf and Gc in adult development is the gulf between the theory of Gc and its measurement. As Cattell (1971) suggested, there are two ways in which Gc can be assessed in adults: One can measure historical Gc — namely the educational and experiential content presumably acquired during adolescence, when individuals are thought to have relatively uniform educational and experiential exposure to the dominant culture — or one can measure Gc directly. The problem with direct measurement of Gc for adults is that there are many different domains of educational and experiential knowledge, such as vocational or avocational knowledge that is unique to either individuals or classes of individuals (e.g., knowledge of law by lawyers, knowledge of child development by day-care workers, and so on). Cattell suggested that with this perspective, an "effort to measure Gc in practice might amount to producing as many tests as there are occupations" (p. 144). To date, investigators have measured only a few domains (e.g., vocabulary, fluency, verbal comprehension). Other intelligence test constructions (such as the Weschler subscales of Information and Comprehension) either fall within the domain of historical Gc or tap knowledge reflecting dominant cultural exposure. Such an approach gives adults virtually no credit for any knowledge that is not common to a wider cultural milieu.

A discussion of the incomplete nature of Gc assessment might be thought of as more of an academic than a practical concern (though decades of controversy about whether adults are more or less intelligent than adolescents defies such a viewpoint). However, numerous developments in other research areas have shifted away from a belief that either machines or people effectively solve most real-world problems with brute-force Gf processes of reasoning and rote memory. Indeed, success in real-world problem solving and decision making has often been attributed to the specific knowledge that is brought to bear on a problem. If performance on real-world intellectual tasks is best predicted by knowledge, emphasizing Gf abilities over Gc is probably a poor choice for comprehensive adult intellectual assessment.

Little is known about the relationships between aging and job performance. One recent meta-analysis of age and job-performance data (McEvoy & Cascio, 1989) illustrates how complex the age-performance situation is. McEvoy and Cascio (1989) found a zero average correlation between age and performance, with a wide range of individual correlations (from r = -.44 to .66). Because any particular job is likely to have differential demands on abilities, it is often difficult to predict when and where age-performance correlations should be found. However, the limited research on the preservation of knowledge across the life

span is more encouraging than the research on fluid abilities. Stanovich, West, and Harrison's (1995) research supported the notion that diverse crystallized knowledge is well-preserved even in 80-year-olds.

Reconceptualizing Adult Intelligence

Various gerontological researchers have attempted to reconcile the fact that adults often perform poorly on several tests of intellectual abilities but nonetheless function quite well in their day-to-day professional and hobby activities (e.g., see the review by Cornelius, 1990). Some researchers have attempted to broaden the treatment of intelligence beyond the IQ test to consider things like practical intelligence (Wagner & Sternberg, 1985) and wisdom (Baltes & Staudinger, 1993). Similar advances in cognitive psychology (e.g., Chi & Ceci, 1987; Ericsson, Krampe, & Tesch-Romer, 1993) have focused on the importance of acquired knowledge in determining the likely success in lifetime scientific productivity (Simonton, 1988) and in context-dependent problem solving, such as physics, chess, and music performance. These researchers have determined that expertise in many fields is predicated on long study and practice in the development of knowledge structures, though some have gone so far as to suggest that traits like intelligence are fundamentally irrelevant for determining individual differences in acquired expertise (e.g., Ericsson et al., 1993).

We focus in this article on a relatively small set of particular ability and non-ability traits for the prediction of individual differences in knowledge. The specific basis for this selection of variables is from two sources: a theory of adult intellectual development (Ackerman, 1996) and investigations that suggest particular personality and interest traits that consistently show overlap with cognitive ability and knowledge measures (Ackerman & Heggestad, 1997; Rolfhus & Ackerman, 1996).

PPIK Theory

An attempt to integrate the perspectives of Hebb (1942) and Cattell (1943) on adult intellect, along with the various results from educational and other areas of psychology, was recently offered by Ackerman (1996, 1998). The framework is broadly developmental, from adolescence through middle-aged adulthood. There are four major components to the theory: intelligence-as-process, personality, interests, and intelligence-as-knowledge (thus, PPIK), and it is illustrated in Figure 1. The theory specifies, as did Cattell, that Gc forms mainly out of Gf, but it differentiates between Gc as it is typically assessed and Gk (intelligence-as-knowledge), which is both broader and deeper than the traditional Gc assessment. It has as a main focus, *typical* intellect, rather than maximal effort (see Ackerman, 1994), in that knowledge can be

Figure 1

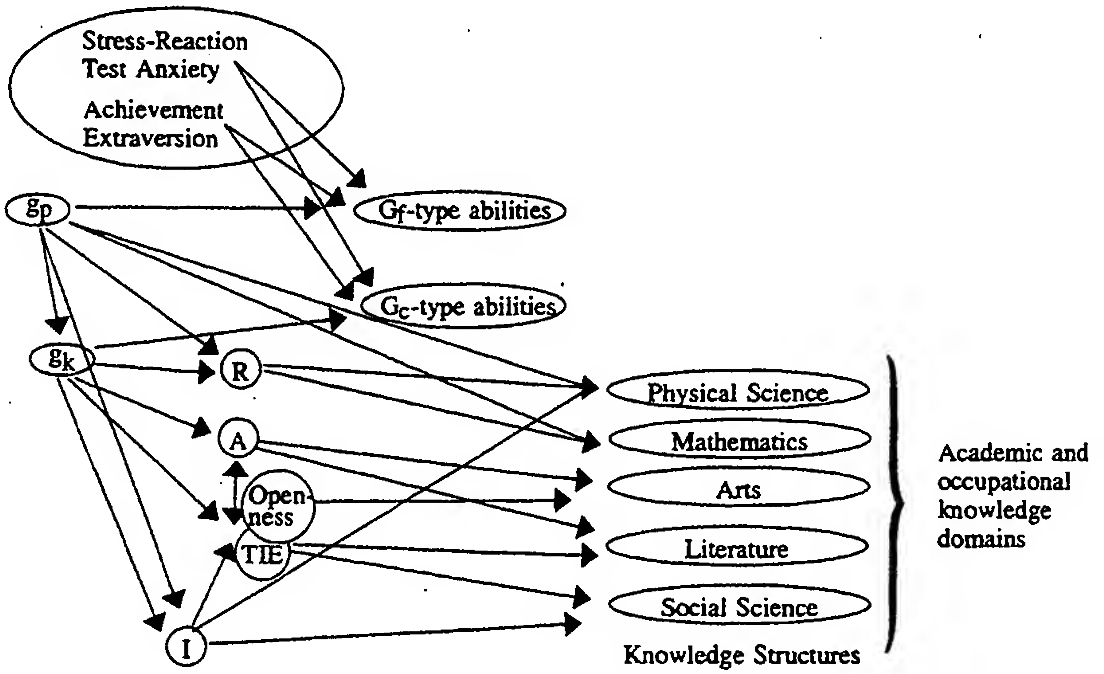


Illustration of Ackerman's (1996) PPIK theory, outlining the influences of intelligence-as-process, personality, interests, and intelligence-as-knowledge during adult development, covering academic and occupational knowledge. Arrows represent correlational influence. The representation indicates that measured fluid intelligence develops out of intelligence-as-process and that measured crystallized intelligence develops out of (or is consequence of) intelligence-as-knowledge. g_c = crystallized; g_f = fluid intelligence; g_p = intelligence-as-process; g_k = intelligence-as-knowledge; R = realistic interests; I = investigative interests; A = artistic interests; TIE = Typical Intellectual Engagement scale. The figure has been adapted and expanded from Ackerman's (1996) article.

accumulated only through effort expended over a long period of time. Moreover, it specifies two sources of personality-ability influences, general and specific. One set of personality traits is hypothesized to have general effects on both intelligence-as-process and intelligence-as-knowledge (achievement, extroversion, stress reaction, and test anxiety). The other set of personality traits (openness to experience and typical intellectual engagement [TIE]) is hypothesized to have positive associations with intelligence-as-knowledge. Three inter-

---

[1] According to Holland (1959), there are six major interest themes, namely realistic, investigative, artistic, social, enterprising, and conventional. Persons who express realistic interests "enjoy activities requiring physical strength, aggressive action, motor coordination and skill" (p. 36). Persons expressing investigative interests are "task-oriented people who generally prefer to 'think through,' rather than 'act out,' problems. They have marked needs to organize and understand the world" (p. 36). Persons who express artistic interests "prefer indirect relations with others. They prefer dealing with environmental problems through self expression in artistic media . . . . They resemble persons with an intellectual orientation in their intraceptiveness and lack of sociability"

est traits are hypothesized to have positive relations to intelligence-as-knowledge: realistic interests[1] associated with knowledge in the areas of physical sciences and mathematics, artistic interests associated with arts and literature, and investigative interests associated with knowledge in the social and physical sciences. Associations among these various traits and knowledge have been found with self-report scales (Rolfhus & Ackerman, 1996).

Personality-Interest-Ability Trait Overlap

The associations among abilities, personality traits, and interests that are described in the PPIK theory have been supported by the previously mentioned review and meta-analysis of Ackerman and Heggestad (1997; see also Ackerman, 1997). From this work, four broad trait complexes were identified, and these are shown in Figure 2. Note that the term trait complex is adapted from Snow (1989). For the current usage, trait complexes are considered to be sets of traits that are sufficiently interrelated to suggest the possibility of mutually causal interdependencies. The four trait complexes are as follows.

1. Social

The social trait complex includes personality traits, such as extroversion and social potency, and interest traits, such as enterprising and social. Persons who are identified with this trait complex are evenly distributed across most abilities.

2. Clerical/Conventional

This trait complex has conventional interests, along with conscientiousness, traditionalism, and control personality constructs. The only salient ability traits associated with this complex are perceptual speed and computational numerical ability. Persons identified with this complex tend to be conventional and "plodding."

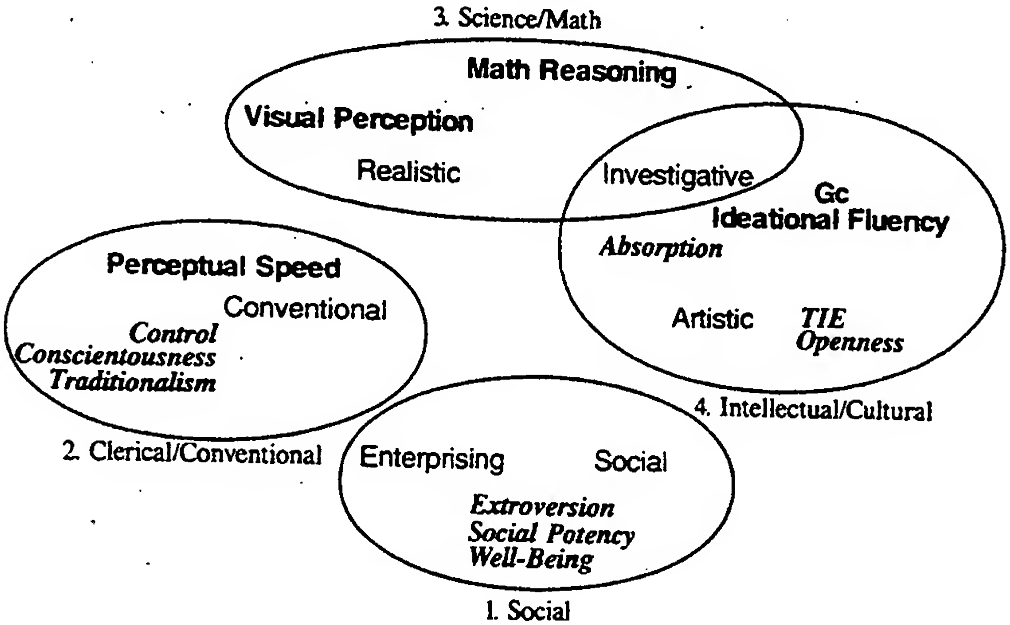
3. Science/Math

This trait complex has substantial overlap with investigative and realistic interests and spatial and mathematical reasoning abilities. In many ways, this

---

(p. 36). Persons who have social interests "prefer teaching or therapeutic roles, which may reflect a desire for attention and socialization in a structured, and therefore safe, setting. They possess verbal and interpersonal skills" (p. 36). Persons who express enterprising interests "prefer to use their verbal skills in situations which provide opportunities for dominating, selling, or leading others . . . .They avoid well-defined language or work situations as well as situations requiring long periods of intellectual effort" (p. 37), and persons who express conventional interests "prefer structured verbal and numerical activities, and subordinate roles. They achieve their goals through conformity" (p. 37).

Figure 2



Trait complexes, based on an analysis of trait correlations through meta-analysis and integrative review. The four trait complexes include abilities, interests, and personality traits showing positive commonalities. Shown are (a) Social, (b) Clerical/Conventional, (c) Science/Math, and (d) Intellectual/Cultural trait complexes. Ability traits are in bold; interests are in standard font; personality traits are in italics.

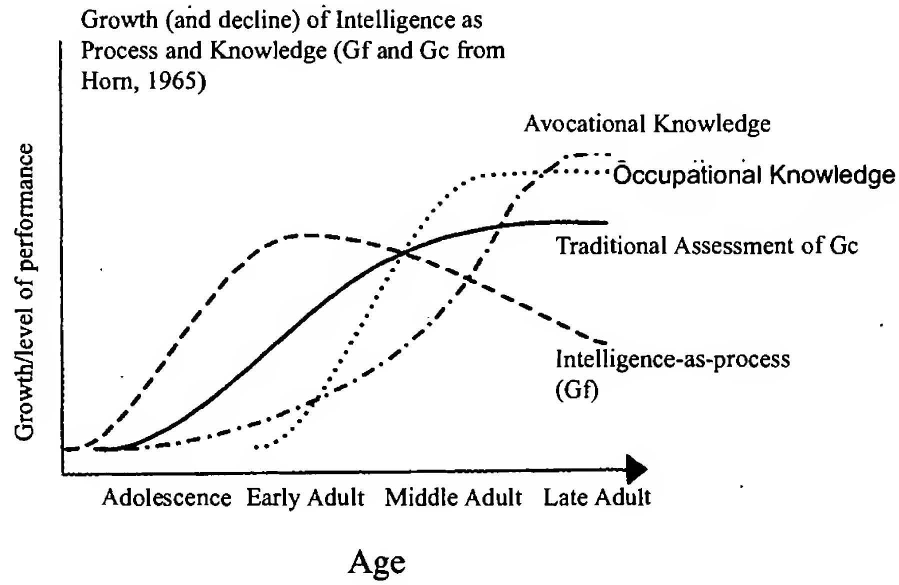
seems to be the trait complex that is most associated with strong Gf, and the interests that correspond are associated with the pursuit of occupations that are most demanding of Gf abilities.

4. Intellectual/Cultural

This trait complex includes investigative interests and artistic interests but also includes the personality traits of TIE, openness to experience, and absorption (the latter is defined as a tendency to become totally engaged in imaginative and ideational thoughts). The ability components of this trait complex include Gc-type measures of verbal ability, fluency, and knowledge and achievement.

Although developmental data are lacking, these four trait complexes provide a basis for examining how intelligence-as-knowledge and intelligence-as-process are aligned with personality and interests. One important implication of these trait associations is that persons may develop coherence across adolescent and adult development, such that changes in one domain (e.g., an increase in artistic interests) may influence changes in intelligence-as-knowledge that could not be readily assessed by a focus on traditional Gf and Gc measures.

Figure 3



Hypothetical growth/level of performance curves across the adult life span for intelligence-as-process, traditional measures of Gc (crystallized intelligence), occupational knowledge, and avocational knowledge. That is, expected levels of occupational and avocational knowledge are shown against reference hypothetical patterns of performance on intelligence-as-process and traditional assessment of crystallized intelligence (Intelligence-as-process [Gf] and Gc are modeled after Horn's (1965, Figure 1, p. 185.)

If intelligence-as-knowledge develops as suggested by the representation shown in Figure 3 (partly based on Horn's, 1965, Figure 1), then the fundamental question is how much weight one should give to Gf-type abilities, traditional Gc-type abilities, and intelligence-as-knowledge in determining adult intelligence. If intelligence-as-knowledge is more predictive than intelligence-as-process measures of important behavioral criteria, and if older adults perform better on intelligence-as-knowledge scales in comparison with younger adults, older adults may indeed be more demonstrably intelligent than younger adults and adolescents.

The Present Investigation

The broad goal of the current research is a transformation of the adult intellectual assessment away from abstract reasoning and high-school knowledge-aptitude measures to direct assessment of adult knowledge structures. To begin though, it is necessary to determine whether knowledge is predicted by or redundant with extant measures of individual differences, and whether trait-knowledge relations are concordant with predictions from the the-

oretical approach taken here (namely PPIK and the four trait complexes). Toward this end, it is necessary to first assess extant abilities, personality, and interest traits. We therefore assessed (a) broad content abilities that constitute fluid and crystallized intelligence (namely, math, spatial, and verbal abilities), (b) critical personality traits (openness and TIE) within a broader "Big-Five" representation of normal personality, and (c) critical interest traits (realistic, investigative, and artistic) within a broader representation of interest traits (Holland, 1973).

In addition, on the basis of other associated research, we decided to examine other traits that operate at the interface between objective measures of aptitude and self-reports of perceived aptitudes that include self-concept (individual beliefs about aptitudes measured on an absolute scale, e.g., "I can express ideas or feelings through writing") and self-ratings of abilities (which are normatively based, e.g., "I am above average on Math Ability"). Such measures may be especially informative for predicting knowledge, in that individuals may act on perceived abilities rather than objective ability levels. These measures are more general than self-efficacy assessments that typically focus only on task-specific situations.

For criteria, we created a set of intelligence-as-knowledge assessment scales to evaluate these predictions. Previous research (Rolfhus & Ackerman, 1996) focused on self-report knowledge assessments, but this clearly was only a first step in the program. The next step was to develop a set of objective measures of knowledge that could be administered to young and middle-aged adults. We started with a set of college-level knowledge tests (from the College-Level Examination Program [CLEP] and the Advanced Placement [AP] series) and added tests of several other domains. For details of the scale refinement and modifications, see Rolfhus (1998) and Rolfhus and Ackerman (2000). It is important to note, though, that these domains of knowledge are but a small sample of the entire population of cultural, vocational, and avocational knowledge. The current study is thus an attempt to describe the relations among abilities, self-concept, interests, personality, and individual differences in the depth and breadth of knowledge.

Specific Hypotheses

Several specific hypotheses can be derived from the extant literature and from the PPIK theory. First and foremost, we predicted that middle-aged adults, in general, would perform better on the various knowledge tests than younger adults (i.e., our traditional college-aged norming group). Second, we predicted that although general intelligence would predict individual differences in knowledge for many areas (especially in the physical sciences — see Figure 1), verbal/Gc abilities would show substantial communality with knowledge scales

(independent of general intelligence), especially in the arts, literature, and social sciences.

On the basis of the PPIK theory and the four trait complexes previously identified, we hypothesized that traits of personality (openness and TIE) would show positive correlations with knowledge in the arts, literature, and social sciences (because of communality with the intellectual/cultural trait complex), but neuroticism and agreeableness were not expected to correlate with knowledge scores (again, because they do not show communality with the derived trait complexes). Extroversion was hypothesized to show small positive correlations with most of the knowledge domains. Traits of interests were predicted to show other systematic correlations, where artistic interests would correlate with knowledge in the arts and humanities (the intellectual/cultural trait complex), realistic interests would correlate with knowledge in the physical sciences and technology (the science/math trait complex), and investigative interests would correlate with knowledge in the physical and social sciences (both the science/math and the intellectual/cultural trait complexes). Other interests, namely social and enterprising interests, were not expected to be correlated with these knowledge domains, because they mainly are associated with a nonintellectual trait complex (i.e., the social trait complex).

Self-concept and self-estimates of ability were expected to largely mirror the correlations with objective ability measures, in line with results of Ackerman, Kanfer, and Goff (1995) and Rolfhus and Ackerman (1996) that indicated relatively close correspondence between academic aptitude self-concepts and objective abilities. Thus, the overall picture expected was that knowledge would reflect a broad conceptualization of Gc, along the lines of Cattell's (1971) views, and intelligence-as-knowledge, as conceptualized by Ackerman (1996).

Method

Participants

Participants were recruited from the campus of the University of Minnesota. Inclusion criteria were age between 30 and 59; normal or corrected-to-normal vision, hearing, and motor coordination; and status as a university student or graduate. The study was completed by 135 participants (42 men and 93 women). The participants' mean age was 40.2 years (SD = 7.2 years). The year that the participants graduated from high school (or completed a general equivalency diploma exam) ranged from 1955 to 1990, with a mean graduation date of 1974 (which corresponded closely with reported participant age: r = -.98). Table 1 provides a breakdown of participant age, highest level of education, work status, and occupation (and also similar information from a young adult norming

Table 1

Demographic Description of the Middle-Aged (MA) Sample and the Young-Adult (YA) Norming Sample

	MA		YA	
Demographic characteristic	**N**	**%**	**N**	**%**
Gender				
Male	42	31.1	49	34.3
Female	93	68.9	94	65.7
Age				
18-27			143	100
30-39	70	51.9		
40-49	52	38.5		
50-59	13	9.6		
Highest Education				
High School (or GED)	10	7.4	140	97.9
Associate (2 years)	22	16.3		21.4
BA or BS	66	48.9	1	0.7
MA or MS	27	20.0		
PhD	7	5.2		
Not reported	3	2.2		
Work status				
Full-time employment	82	60.7	6	4.2
Part-time employment	33	24.4	82	57.3
Not employed	12	8.9	54	37.8
Not reported	8	5.9	1	0.7
Occupational classification				
Managerial or professional	57	42.2	6	4.2
Technical, sales, or administrative support	37	27.4	24	16.8
Service	21	15.6	48	33.6
Precision production, craft or repair	3	2.2	0	0.0
Operator, fabricator, or laborer	3	2.2	8	5.6
Military	1	0.7	2	1.4
Not reported (or not employed)	13	9.6	55	38.5

Note. GED = general equivalency diploma.

sample; see Rolfhus, 1998; Rolfhus & Ackerman, 2000). To assess the relationship between age of respondent and highest level of completed education, age group was coded into three levels (ages 30-39, 40-49, and 50-59), and education was coded into three levels (less than a BA, BA only, a degree attained past the BA). We found no significant interaction between age and level of education for the middle-aged group, $X^2(4) = 3.14$, $p = .53$.

Apparatus

Ability tests were administered in a group setting. Tests were compiled into booklets, and the participants wrote answers directly on the tests. Instructions and start-stop timings were presented over a public address system, using prerecorded minidiscs. Self-report items were administered on IBM-compatible computers, and each participant sat at an individual workstation. Knowledge test items were administered with large-font bitmapped graphics on IBM-compatible computers. Responses were made using standard computer keyboards.

Measures

Ability Battery

Tests were selected to assess the following abilities. There were three verbal tests: (a) Verbal Analogies (from Ackerman & Kanfer, 1993), (b) Controlled Associations (from the Educational Testing Service [ETS] Kit; Ekstrom, French, Harman, & Dermen, 1976), and (c) the Extended Range Vocabulary Test (from the ETS Kit). There were three numerical tests: (a) Math Knowledge (from Ackerman & Kanfer, 1993), (b) Problem Solving (from Ackerman & Kanfer, 1993), and (c) Number Series (from the Primary Mental Abilities [PMA] battery; Thurstone, 1962). There were three spatial abilities tests: (a) Paper Folding, (b) the Verbal Test of Spatial Abilities, and (c) Spatial Orientation (all three tests from Ackerman & Kanfer, 1993). There was also one test of mechanical knowledge: Mechanical Knowledge from the Cognitive Ability Battery (Hakstian & Cattell, 1976).

Nonability Traits

Table 2 presents information about the number of items, response format, and internal consistency reliability for each of the nonability scales as follows.

Personality. The NEO-FFI (FFI stands for Five-Factor Inventory; Costa & McCrae, 1992) was included to represent broad personality markers. This inventory is composed of 60 items to measure five factors: Neuroticism, Extroversion, Openness, Conscientiousness, and Agreeableness. The 59-item

Table 2
Characteristics of Nonability Measures

Scale	Total no. of items	M	SD	α
Personality				
Neuroticism	12	32.61	8.92	.88
Extroversion	12	41.04	7.02	.82
Openness	12	46.32	5.66	.68
Agreeableness	12	45.78	6.58	.73
Conscientiousness	12	45.56	6.58	.83
TIE	59	264.09	31.39	.92
Interests				
Realistic	15	54.24	11.99	.85
Artistic	15	68.42	14.01	.90
Investigative	15	66.58	13.08	.91
Social	15	70.10	10.52	.84
Enterprising	15	57.81	13.38	.87
Conventional	15	44.64	15.66	.93
Self-concept				
Mechanical	5	18.75	6.54	.92
Math	5	15.82	7.19	.95
Verbal	5	26.10	3.16	.84
Spatial	5	20.19	5.38	.90
Self-estimates of ability				
Verbal	4	23.18	3.47	.88
Math	4	15.06	4.87	.80
Mechanical	2	8.32	2.36	.14

Note. Responses were scored on the following scales: Personality (NEO Five Factor Inventory) (1 = strongly disagree to 5 = strongly agree); Typical Intellectual Engagement (TIE) scale (1 = strongly disagree to 6 = strongly agree); interests (1 = strongly dislike to 6 = strongly like); self-concept (1 = strongly disagree to 6 = stronly agree); self-estimates of ability (1 = extremely low to 7 = extremely high). Negatively worded items were reverse scored by exchanging, for example, a score of 1 for a score of 6, 2 for 5, 3 for 4, and so on. Means, standard deviations, and internal consistency reliability (Cronbach's alpha) were computed on the middle-aged sample (N = 135).

est themes (factors) identified by Holland (1973), including the following: Realistic (e.g., bookkeeping), Investigative (e.g., studying physics), Artistic inter (e.g., composing music), Social (e.g., running focus groups), Enterprising (e.g., entertaining others), and Conventional (e.g., repairing computers).

TIE scale (Goff & Ackerman, 1992) was administered. Sample items are "There are few topics that bore me" and "I read a great deal." Other scales were part of a larger study and are not reported here.

Interests. The 90-item Unisex Edition of the American College Testing Interest Inventory (UNIACT; Lamb & Prediger, 1981) was used to assess

Self-concept. Self-concepts for competencies and aptitudes were assessed with a 30-item self-report questionnaire (Ackerman et al., 1995; Kanfer, Ackerman, & Heggestad, 1996). The areas assessed were self-concepts of mechanical, verbal, and spatial aptitude competencies, as well as others not reported here.

Self-estimates of ability. For self-ratings of ability, a 21-item questionnaire was used (Ackerman et al., 1995; Kanfer et al., 1996). The items represent five scales of broadly described abilities, aptitudes, or skills, namely verbal, math, mechanical (and others not reported here). Participants responded with a self-evaluation relative to other persons their age.

Participants also completed a take-home questionnaire. To save space, we report only educational background, work status, and occupational classification information from this questionnaire.

Knowledge Scales

College Board examinations were used to form the backbone of the knowledge tests. However, the tests were required to discriminate in populations with less knowledge of these domains than the CLEP and AP test populations because the CLEP and AP tests assess knowledge equivalent to completion of one or two college courses. Easier items were developed to augment the CLEP and AP exams. Prior to final use, all of these tests had been put through at least two development cycles (where all items were administered to at least 100 examinees, the test was revised, and it was then readministered to another 100 or more examinees; see Rolfhus, 1998). The final knowledge scales were con-

[2]Prior to the current study, the knowledge scales were developed through a series of pilot studies (to obtain the necessary item-difficulty statistics that allowed the current tests to be administered in a power-test format). Each domain item pool was administered to at least 100 undergraduate students. After a knowledge scale was administered, the distribution of item difficulties was examined. Deviations from a uniform distribution were identified. New items were written to fill any gaps within the distribution. When multiple items indicated equivalent difficulty levels, items were deleted so that only single items remained at each level. Items that exhibited large gender differences were removed or edited. When new items were written or old items changed, that knowledge scale was readministered in a new sample. Some knowledge scales underwent three rounds of administration and revision. To complete the knowledge tests, five validation studies were run over a period of 18 months. Across the five studies, 700 participants completed at least two hours of testing each, and each participant completed roughly six tests. Extensive details of these development studies can be found in Rolfhus (1998). Internal consistency reliabilities presented in Table 2 were derived from the last pilot study for each scale, when all items in each test were administered.

structed with a power format where the easiest items were administered first, followed by items of increasing difficulty. As in many other power tests, when the participant answered three items in a row incorrectly, that particular test was terminated, and another test was started.

Internal consistency reliabilities (Cronbach's alphas) for the tests ranged from .71 to .94 in earlier pilot studies.[2] The minimum number of items in a test was 35 (Art), and the maximum was 123 (American History). The content of the tests is briefly described below:

American Government: This test covers the structure of American government, the function of various government units, and the American political system.

American History: This test covers a period from prerevolutionary times to the present.

American Literature: This test covers a range of American writers, playwrights, and poets.

Art: This test requires identification and interpretation of art and architecture from around the world from images displayed on the computer monitor.

Astronomy: This test covers observational tools and techniques, the structure of the solar system and the universe, and the physical principles that govern astronomical observations.

Biology: This test covers a broad range at the cellular, organismal, and ecological levels.

Business/Management: This test covers business and management principles and their applications.

Chemistry: This test covers a range of information from atomic structure to standard laboratory procedures.

Economics: This test covers both micro- and macroeconomics.

Electronics: This test covers basic principles of electricity and their applications.

Geography: This test covers geography, including mountains, rivers, oceans, and so on.

Law: This test covers basic principles of law, as well as criminal, civil, and business law.

Music: This test covers basic music terminology and styles, instruments, and composers.

Physics: This test covers basic physical principles and their applications.

Psychology: This test covers the content from an introductory college course in psychology.

Statistics: This test covers the content of an introductory college course in statistics.

Technology: This test assesses understanding of a wide range of modern technologies.

Tools/Shop: This test covers both tool identification and tool use.

Western Civilization: This test covers major political, philosophical, and eco-

nomic events in Europe from Ancient Greece to the Cold War.

World Literature: This test covers mainly classic Western literature and poetry.

Procedure

The study was conducted in two 3-hour sessions, supplemented by a take-home biographical inventory. The first session was devoted to the assessment of traits. After completion of the consent form, participants were given ability tests, which consisted of the three tests each for verbal, spatial, and numerical abilities, along with the test of mechanical reasoning. After completion of the ability tests, participants moved to the computers for assessment of the nonability traits (including the NEO-FFI, TIE, UNIACT) and measurement of self-concept and self-estimates of ability. At the end of the first session, participants were given a take-home questionnaire. The second session was devoted to the assessment of knowledge in 20 different areas. Domains assessed represented four previously determined knowledge factors, including Humanities, Sciences, Civics, and Mechanical (Rolfhus, 1998).[3]

Results

Several sets of results are presented here: (a) a description of how the participants performed on the ability and knowledge measures, in comparison with the norming sample (reported in Rolfhus & Ackerman, 2000), (b) education-ability and education-knowledge correlations, (c) an analysis of the ability-knowledge correlates and their underlying factor structures, (d) nonability correlates of knowledge, and (e) a series of multiple correlation assessments of the incremental validity of demographic, ability, personality, and interest variables for predicting knowledge-scale scores.

Comparison With the Young-Adult Norming Group

All of the measures administered in this study have also been previously or concurrently administered to young-adult samples. For cross-sectional comparison purposes, we review the performance of the current sample in comparison with a concurrent sample of students from an introductory psychology course (N = 143; ages 18-27, $M = 19.1$, $SD = 1.2$). Additional information on this sample can be found in Rolfhus and Ackerman (2000) and Rolfhus (1998). Although these two samples are not identical in population, in selection (the

[3]The categorization of knowledge scales into Humanities, Sciences, Civics, and Mechanical factors was derived through a variety of different exploratory techniques, namely rational categorization, cluster analysis, multidimensional scaling, and factor analysis. The final categorization emerged from these different techniques. A different sample of either scales or participants might yield a different categorization.

young adults received course credit and monetary remuneration, and the current sample received only monetary remuneration), or in education (only one of the young adults in the sample had completed a BA level of education), a comparison between the two groups provides a baseline of ability and knowledge levels that allows us to evaluate the nature of the current sample. The first set of analyses provides an overview of the similarities and differences among the norming group of young adults (the YA group), the current sample of middle-aged adults (the MA group), and to some degree, within the MA group across the age range. For each measure, means and standard deviations are presented for the YA and MA groups, followed by the results of *t* tests for the differences between means. In this way, it is possible to look at broad differences between the YA and MA samples (which admittedly come from two different populations — one sample of students from an introductory psychology course and the other sample from the university community at large). Finally, within the MA sample, it is possible to examine the relationship between age and performance (across the 29-year age range from youngest to oldest examinee in the MA group).

Abilities

A standard cross-sectional comparison between the MA and YA groups was performed to evaluate the general representativeness of the sample. On the basis of extant literature (e.g., Horn, 1965), a priori expectations were that the MA group would show higher scores on the verbal tests and equivalent or lower scores on the numerical and spatial tests. Two sets of comparisons were computed: mean differences between the age 18-27 and age 30-59 groups, and within-group correlations between test scores and chronological age and between test scores and education for the age 30-59 participants. These results are shown in Table 3.

In reviewing the means, standard deviations, and *t* test results, it is clear that the MA sample performed substantially better than the YA group on the verbal ability tests and the Mechanical Knowledge test, performed less well on the numerical ability tests (except for the word-problem test), and had mixed results on the spatial ability tests. Table 3 shows differences in raw scores. We also calculated what the mean levels of performance for the MA group would have been (in z scores) if the ability tests had been standardized to the young examinees (Verbal Analogies = 1.0, Controlled Associations = 0.83, Extended Range Vocabulary Test = 2.44, Paper Folding = -0.45, Verbal Test of Spatial Abilities = 0.14, Spatial Orientation = 0.00, Math Knowledge = -1.2, Problem Solving = 0.11, Number Series = -0.49, Mechanical Knowledge = 0.48). With unit-weighted z score composites for three major ability factors, the aggregate mean scores for the MA group (again with the YA sample providing the standardization; see Rolfhus, 1998; Rolfhus & Ackerman, 2000) were also computed (Verbal Composite z = 1.6, Numerical Composite z = -0.67, and Spatial Composite z =

Table 3

Characteristics of Reference Aptitude and Ability Tests for the Young-Adult (YA) and Middle-Aged (MA) Groups

			YA		MA			MA (only)	
Ability Test	No. of items	a	M	SD	M	SD	*t* test	*r* with age	*r* with education
Verbal Ability									
1. Verbal Analogies	50	.68	26.71	5.75	32.68	5.58	-8.79***	-.01	.33**
2. Controlled Associations	8	.78[a]	23.24	7.67	29.61	9.47	-6.19***	-.01	.35**
3. Extended Range Vocabulary Test	48	.86	15.19	6.72	31.62	7.58	-19.15***	.15	.29**
Numerical ability									
4. Math Knowledge	32	.78	17.35	6.36	9.69	5.69	10.55***	-.21*	.24**
5. Problem Solving	15	.36	4.30	2.29	4.56	2.19	-0.99	-.11	.15
6. Number Series	20	.72	10.22	2.84	8.81	3.21	3.88***	-.18*	.12
Spatial ability									
7. Paper Folding	24	.73	12.39	5.06	10.09	4.87	3.87***	-.28**	.15
8. Verbal Test of Spatial Abilities	24	.72	11.49	4.34	12.08	4.90	-1.06	-.19*	.26**
9. Spatial Orientation	20	.59	7.18	3.91	7.21	3.83	-0.05	-.30**	.11
Mechanical knowledge									
10. Mechanical Knowledge	18	.69	6.96	3.95	8.87	3.83	-4.10***	.01	.01

Note. The YA group ranged from 18 to 27 years of age; the MA group ranged from 30 to 59 years of age. N = 143 (YA, from Rolfhus & Ackerman, 2000) and N = 135 (MA); for the *t* test between the YA norming sample and the MA group, *df* = 276; for the correlations, df = 133.

[a]The reliability index is based on a Part I-Part II correlation and the Spearman-Brown prophecy formula. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 4

Characteristics of the Knowledge Scales (Number Correct) for the Young-Adult (YA) Norming Sample and the Middle-Aged (MA) Group

	YA		MA			MA (only)	
Knowledge scale	M	SD	M	SD	*t* test	*r* with age	*r* with education
Humanities							
1. American Literature	33.90	14.61	55.83	16.49	-11.70***	.11	.35**
2. Art	12.13	6.03	16.41	4.74	-9.82***	-.09	.18
3. Geography	20.81	12.20	29.52	13.35	-5.27***	-.06	.08
4. Music	19.88	10.15	33.44	10.25	-10.34***	.13	.28**
5. World Literature	30.79	14.18	53.80	18.74	-11.55***	.05	.31**
Sciences							
6. Biology	16.22	10.83	23.31	12.27	-4.98***	-.09	.19*
7. Business/Management	13.20	7.65	23.67	9.26	-10.27***	.00	.26**
8. Chemistry	15.93	8.86	15.88	9.41	0.05	-.12	.26**
9. Economics	17.97	13.44	27.99	14.02	-6.01***	-.11	.32**
10. Physics	12.82	6.46	15.91	5.84	-4.13***	-.06	.35**
11. Psychology	13.94	7.56	19.71	7.98	-6.18***	.01	.17
12. Statistics	7.44	4.55	8.78	5.50	-2.10*	-.13	.34**
13. Technology	18.82	11.07	30.68	12.58	-7.52***	-.03	.17
Civics							
14. American Government	24.14	13.63	39.66	13.24	-9.47***	-.03	.28**
15. American History	38.88	20.30	61.80	20.63	-8.51***	.03	.39**
16. Law	17.76	10.40	30.95	10.14	-10.62***	-.06	.22*
17. Western Civilization	22.82	12.62	34.94	15.49	-6.90***	.04	.26**
Mechanical							
18. Astronomy	16.34	10.11	19.69	12.73	-2.23*	.17	.14
19. Electronics	10.52	5.92	18.16	7.56	-8.70***	.14	-.06

20. Tools/Shop	16.11	6.92	20.98	7.40	-5.72***	.19*	-.01

Note. The YA group ranged from 18 to 27 years of age; the MA group ranged from 30 to 59 years of age; the number of participants varied (because of incomplete data): max N = 143 (YA) and 135 (MA); for the t test between the YA norming sample and the MA group, df = 276; for the correlations, max df = 133.

*$p < .05$. **$p < .01$. ***$p < .001$.

-0.13). Such results are generally consistent with the extant literature because the MA group showed lower numerical ability and higher verbal ability. However, in keeping with our sampling strategy of adults throughout the university community (many of the examinees were students in the Continuing Education and Extension programs), the MA examinee group appears to have had somewhat higher abilities than might otherwise have been found in the at-large population (which is also concordant with the higher proportion of examinees having completed BA level post-secondary education than would be found in the wider community).

The MA examinee within-group correlation between age and performance provided further evidence in support of the general representativeness of the MA adults, at least with respect to ability levels and age/cohort differences. The correlations between age and test score, shown in the last column of Table 3, indicated significant negative relations between age and ability for all three spatial ability tests and for two of the numerical ability tests (excluding Problem Solving, which contains math-type word problems and has a substantial verbal component), and indicated no significant correlations between age and ability on the verbal ability tests and the Mechanical Knowledge test. Level of education also correlated positively and significantly with five ability test scores, including all three verbal tests and one test each from the math and spatial domains. Because age and education were essentially uncorrelated in the MA sample (r = -.03), the associations between age and test performance and between education and test performance are essentially independent contributions to explaining test performance.

Knowledge Scales

Similar to the analysis of ability tests, the analysis of knowledge scales first addressed differences between the current MA sample and the YA norming sample for the knowledge scales and, for the MA group, addressed correlations between knowledge and

age. A set of analyses was conducted for the knowledge scales that paralleled the earlier analysis of traits (see Table 4). In reviewing the results shown in Table 4, note that it is important to refer back to Table 3. That is, Table 3 shows that the MA group, although performing better than the YA group on the verbal abilities, performed more poorly on the numerical abilities — as would be expected. However, in all but one of the 20 knowledge tests shown in Table 3 (the Chemistry test), the current sample (ages 30-59) performed significantly better than the age 18-27 norming group. In addition, only one of the tests showed a significant correlation with age for the MA group (Tools/Shop showed an increase with age), suggesting that factors other than age are primarily responsible for determining individual differences in the depth and breadth of knowledge, at least within a broad range of middle-adult years. Although the sample is not sufficiently large to provide substantial power to decisively retain a hypothesis of zero correlations, the lack of significant negative correlations is entirely consistent with an intelligence-as-knowledge perspective.

Level of education was positively and significantly correlated with knowledge-scale performance in each of the domains except for the mechanical area. Again, because age and education were essentially uncorrelated in the MA sample, level of education provides an independent contribution to explaining individual differences in amount of knowledge in many areas (but clearly not all of the areas).

Although the MA group performed better overall on the knowledge scales, the MA group did not have equivalently higher scores throughout the broad range of tests. With the taxonomic categorization of knowledge domains, knowledge-domain *z* score composites can be formed to provide a more broad-brush representation of the YA-MA group differences. As with the ability scales, the knowledge-scale scores were standardized on the YA norming group, and then mean *z* scores were computed for the MA group. We calculated group differences, and the areas of greatest advantage to the MA group were, in descending order, as follows: For the humanities the mean *z* score difference was 1.28 *SD* units; for civics the mean *z* score difference was 1.12 *SD* units; for mechanical the mean *z* score difference was 0.80 *SD* units; and for the sciences the mean *z* score difference was 0.67 *SD* units.

These results provide additional support for the sense that the MA group had higher mean performance in all of the knowledge categories that were assessed in this study — though the largest differences between the MA and YA groups were found in the Humanities and in Civics, and the smallest differences in knowledge were in the sciences. One potential interpretation here is that the MA sample excelled in domains that are traditionally found in programs offered to nontraditional adult students and that the YA norming sample came closest to holding even with the MA group on the physical science topics (such as chemistry and physics). In fact, in any given academic quarter, fewer continuing education courses are offered at the University of Minnesota in physics (7) and

chemistry (4) than in art (45) or literature (19). For the 12 domains with clearly identified university courses in the fall 1997 course catalog, the correlation between the magnitude of MA advantage over YA group and number of courses offered was .66 ($p < .05$); that is, greater MA advantages were found for topics with more numerous course offerings.

Education

With the coding of highest level of education as an ordinal variable (0 = high school or equivalent, 1 = BA/BS or equivalent, and 2 = post-BA/BS education), we computed correlations with abilities and knowledge scales. With the exception of the mechanical knowledge scales, most of the correlations between education and knowledge were significant and positive — especially in the sciences and civics domains. As might be expected, positive correlations were found between highest level of education and ability (.39, .22, and .21 for verbal, spatial, and numerical composites, respectively). In contrast, a dichotomous variable for current student status (0 = not a current student, 1 = current student) yielded only two significant correlations with knowledge (Biology = .21 and Statistics = .26) and essentially zero correlations with ability.

Ability-Knowledge Correlations

To many researchers interested in adult intelligence, the most important results of this investigation will be captured in the associations between traditional ability factors and knowledge-scale performance. Although a review of individual ability-test/knowledge-scale correlations is illuminating, an attempt has been made to summarize the noteworthy associations between the battery of ability tests and the battery of knowledge scales. There are many statistical techniques that can be used to describe the communality between these two sets of scales, such as interbattery factor analysis and canonical correlation analysis. The first pass through these data is less complicated (and arguably less sophisticated), but it is a method that is easy to understand (at least conceptually). The method adopted here starts with a hierarchical factor analysis of the ability tests (using the procedure described by Schmid & Leiman, 1957), so that an oblique first-order factor solution can be recast as an orthogonalized factor matrix of two orders (in this case, a general ability factor, *g*, that has substantial loadings across all three major ability factors: Verbal, Spatial, and Numerical). With a derivation of the factors underlying the ability tests, factor loadings of the knowledge scales are estimated by use of Dwyer's (1937) extension procedure. Dwyer's procedure (and the generalization by Mosier, 1938; see Gorsuch, 1983, for a more recent description) is a general linear model approach that allows one to correlate factors derived from one set of variables with new variables, without using factor scores.

Table 5

Ability Factor Hierarchical Factor Solution and Dwyer-Extended Loadings to Knowledge Scales

Scale	*g*	Verbal	Spatial	Numerical	h^2
		Hierarchical factor solution			
Verbal Analogies	**.634**	**.610**	-.001	.179	.806
Extended Range Vocabulary Test	**.377**	**.694**	.016	-.010	.624
Controlled Associations	**.356**	**.436**	-.013	.076	.322
Math Knowledge	**.603**	.114	.161	.238	.460
Problem Solving	**.578**	.008	.243	.216	.440
Number Series	**.631**	-.048	.008	**.385**	.549
Paper Folding	**.437**	.040	**.468**	.007	.412
Verbal Test of Spatial Abilities	**.607**	.026	**.482**	.105	.613
Spatial Orientation	**.385**	-.089	**.517**	-.007	.423
		Ability factor: Knowledge-scale correlations			
Humanities					
American Literature	.156	**.624**	.016	-.117	
Art	.210	**.552**	**.378**	-.246	
Geography	**.419**	.149	.197	.100	
Music	.272	**.504**	-.029	.013	
World Literature	.227	**.639**	.080	-.113	
Sciences					
Biology	**.454**	**.439**	.240	.004	
Business/ Management	.281	**.364**	.185	-.045	
Chemistry	**.567**	.238	.134	.190	
Economics	**.414**	.152	.117	.136	
Physics	**.557**	**.366**	.240	.089	
Psyhcology	.253	**.449**	.097	-.045	
Statistics	**.431**	.217	.112	.128	
Technology	**.500**	**.530**	.174	.036	
Civics					
American Govt.	**.462**	**.303**	.140	.104	
American History	**.371**	**.430**	.077	.041	
Law	**.377**	**.302**	.209	.019	
Western Civilization	**.320**	**.516**	.034	.006	
Mechanical					
Astronomy	**.366**	**.350**	.141	.031	
Electronics	**.438**	.188	**.340**	.025	
Tools/Shop	**.367**	.200	**.461**	-.082	

Note. Salient loadings (defined a loadings ≥ .300) are shown in bold. *g* = general intelligence.

The hierarchical factor solution of the ability measures and Dwyer-extended factor loadings for the knowledge scales are shown in Table 5. Note that for the factor analysis of abilities, the Mechanical Knowledge test was not included, because as a single ability test, it could not be used to sufficiently identify a factor. In addition, the construct underlying the test (mechanical reasoning or mechanical knowledge) falls between the ability and knowledge domains. The ability factor solution provided no surprises; the intercorrelations among the nine tests showed positive manifold, and three first-order factors were easily identified as Verbal, Spatial, and Numerical Abilities. Rotation to an oblique simple-structure criterion (the Tucker-Finkbeiner direct personal artificial probability factor rotation method; see Tucker & Finkbeiner, 1981) indicated positive correlations among the three factors. The Schmid-Leiman orthogonalized hierarchical factor solution shown in Table 5 indicates that all of the nine ability tests significantly loaded on the second-order *g* factor, with salient first-order factor loadings for the verbal and spatial tests and somewhat smaller loadings of the numerical tests.

The Dwyer extension results showed that first of all, after accounting for a general ability factor (*g*), very little common variance remained between Numerical or Spatial Ability factors and knowledge scales. Only the Art, Electronics, and Tools/Shop tests had salient loadings on the Spatial Ability factor, and none of the knowledge scales had a salient loading on the Numerical Ability factor. In contrast, many of the knowledge scales showed substantial positive loadings on the first-order Verbal Ability factor (which is as close as traditional ability assessment methods come to assessing intelligence-as-knowledge in adults), as predicted by the PPIK theory. By and large, the largest loadings on the Verbal Ability factor came from knowledge scales in Humanities, followed by Sciences and Civics. The second-order *g* factor was significantly and positively associated with 14 of the 20 knowledge scales. The humanities scales notably showed the overall lowest loadings on the *g* factor.

In summary, the factor solution and the extension analysis indicated that an underlying Verbal Ability factor showed substantial common variance with many (14) of the knowledge scales, even in addition to the variance in common between the verbal tests and *g*. In contrast, Numerical and Spatial Abilities (which figure prominently in traditional measures of intelligence) appear to have substantially less common variance with knowledge in the domains under consideration, once their common variance associated with general intelligence is partialled out. As such, the PPIK inspired hypotheses are supported. That is, verbal/crystallized abilities are more highly associated with what adults know across a wide variety of topics, in contrast to the communality between abilities that are more associated with intelligence-as-process (i.e., spatial and numerical abilities) and what adults know.

The demonstration of an ability-knowledge association is but a modest effort in validating the PPIK theory and in putting the theory to use in educational and

occupational applications. The next step is to take an account of other trait correlates of individual differences in knowledge. In the remaining analyses, we attempted to further validate the nonability portions of the PPIK theory and to assess the incremental validity of the various predictors.

Nonability Correlates of Knowledge

Self-Concept and Self-Estimates of Ability

Correlations between self-report measures of self-concept and self-estimates of ability with the knowledge scales are reported in Table 6. Consistent with earlier investigations that correlated self-concept scales and abilities, self-report knowledge, and other traits (e.g., see Ackerman, 1997; Ackerman et al., 1995; Rolfhus & Ackerman, 1996), there is thematic correspondence between self-concept, self-estimates of abilities, and objective data (e.g., objective aptitude measures). In this case, self-concepts of mechanical ability or skill and self-estimates of mechanical knowledge correlated positively and significantly with individual differences in scores on the sciences knowledge scales (e.g., Physics and Chemistry) and with the mechanical knowledge scales (i.e., Tools/Shop, Electronics, and Astronomy). Similarly, math self-concept and self-estimates of math ability correlated with sciences knowledge (and mechanical knowledge scales). Spatial aptitude self-concept correlated significantly and saliently with scores on the Geography, Technology, Astronomy, Electronics, and Tools/Shop tests.

In a noteworthy but not surprising finding, individual differences in math self-concept were not significantly associated with most of the humanities and civics knowledge scales, and verbal self-concept was not significantly related to many of the sciences and mechanical knowledge scales. One reason for this was the essentially zero correlation between math self-concept and verbal self-concept (.02), even though objective math and verbal ability scores were positively correlated (.46). For a discussion, see Ackerman et al. (1995).

Verbal self-concept and self-estimates of verbal ability showed substantial positive correlations with many of the knowledge scales and showed negative (but not significantly negative) correlations with only Economics and Tools/Shop. The largest communalities were found for humanities and civics. As expected, smaller communalities were found for mechanical knowledge scales and sciences knowledge scales. Such results are again consistent with the predictions made by the PPIK theory, supporting a broad framework for crystallized intelligence, both in aptitude and in nonability traits, such as self-concept.

Interests-Knowledge

Table 7 shows the correlations between interest theme scores from the UNI-ACT and knowledge-scale scores. Correlations between interests and knowledge

Table 6

Self-Estimates of Ability and Self-Concept Correlations with Knowledge

	Verbal		Math		Mechanical		Spatial
Knowledge scale	SE	SC	SE	SC	SE	SC	SC
Humanities							
American Literature	**.418*** **	**.303** **	.057	.000	-.008	.052	.051
Art	.284**	.377**	.065	.085	-.054	.043	.142
Geography	.094	.106	**.328** **	.244*	.101	.201*	**.385** **
Music	.204*	.257**	.219*	.148	.105	.119	.090
World Literature	**.370** **	**.322** **	.065	.028	-.060	.035	.069
Sciences							
Biology	.256**	.264**	**.313** **	.224*	.088	.254**	.186*
Business/Management	.150	.084	.149	.166	.093	.156	.134
Chemistry	.012	.072	**.550** **	**.519** **	.148	**.379** **	.235**
Economics	-.003	-.032	**.414** **	**.418** **	.110	.233**	.211*
Physics	.058	.098	**.443** **	**.380** **	**.312** **	**.421** **	.213*
Statistics	.146	.039	.194*	.155	.056	.145	.063
Technology	.170	.179	**.382** **	.262*	.196	.295**	**.300** **
Civics							
American Government	.144	.039	**.326** **	.245**	.046	.080	.037
American History	.203*	.081	.180	.145	.160	.214*	.132
Law	.184*	.178*	.231**	.202*	.097	.146	.179*
Western Civilization	**.344** **	.185*	.200*	.094	-.020	.075	.142
Mechanical							
Astronomy	.102	.048	.283**	.184	.166	**.327** **	**.326** **
Electronics	.067	.133	**.455** **	**.356** **	.279**	**.510** **	**.430** **
Tools/Shop	-.058	-.008	.295**	.245**	**.411** **	**.512** **	**.320** **

Note. Salient correlations ($\geq$ .300) are shown in bold. SE = self-estimate; SC = self-concept. *$p < .05$. **$p < .01$.

Table 7
Interest-Knowledge Correlations

Knowledge scale	Realistic	Investigative	Artistic	Social	Enterprising	Conventional
Humanities						
American Literature	-.042	.157	.177	-.057	-.083	-.080
Art	-.004	.164	**.361****	-.074	-.128	-.181
Geography	.111	.174	-.040	-.084	-.118	.045
Music	.067	.174	.123	-.047	-.144	-.021
World Literature	-.017	.175*	.184*	-.050	-.108	-.024
Sciences						
Biology	.189*	**.368****	.001	-.046	-.190	-.171
Business/Management	.054	.098	-.078	-.062	.037	.045
Chemistry	.158	**.344****	-.196*	-.086	-.158	-.013
Economics	.113	.074	-.157	-.102	-.027	.100
Physics	.239**	**.327****	-.050	-.038	-.126	-.031
Psychology	.054	.229**	.180*	.171*	.030	.023
Statistics	.046	.078	.161	.048	-.053	-.141
Technology	.165	**.352****	-.016	-.196	-.271**	-.012
Civics						
American Government	-.013	.028	-.053	-.151	-.139	-.035
American History	.086	.171	.003	-.101	-.080	-.030
Law	.090	.056	-.120	-.171*	-.029	.113
Western Civilization	-.014	.191*	.089	.024	-.061	-.066
Mechanical						
Astronomoy	.179	**.322****	-.015	-.133	-.144	.020
Electronics	**.370****	**.336****	-.004	-.045	-.121	.066
Tools/Shop	**.418****	.182*	-.144	-.060	-.132	.071

Note. Salient correlations (≥ .300) are shown in bold. *$p < .05$. **$p < .01$.

scales are concordant with the predictions from the PPIK theory. That is, the domains of knowledge under consideration were generally not associated with individual differences in Social, Enterprising, or Conventional interests but were positively and significantly associated with Realistic, Investigative, and Artistic interests. A salient positive correlation was found between Artistic interests and knowledge about art.

Similarly, Realistic interests showed positive correlations with knowledge scales in the mechanical domain (viz., Electronics and Tools/Shop). Finally, Investigative interests, correlated positively and significantly with 10 of the knowledge scales, most notably in the sciences and the mechanical knowledge domains. Across these domains, there was a significant correspondence between Realistic, Investigative, and Artistic interests and actual level of knowledge acquired. The question of where to put the "causal arrow" (i.e., interests ® knowledge, knowledge ® interests, or some other variable ® both interests and knowledge) is an enduring issue for future research (e.g., see Sorenson, 1933, 1938).

Personality-Knowledge

The personality trait correlations with knowledge-scale scores shown in Table 8 provide general confirmation of predictions. That is, the Neuroticism and Agreeableness factors of the NEO-FFI do not correlate significantly with the knowledge scales. The TIE scale and the Openness factor of the NEO-FFI do correlate highly and significantly with the knowledge scales, especially in the humanities and civics domains. Interestingly, the Conscientiousness factor fails to correlate significantly positively with any of the knowledge scales (and correlates negatively with the humanities and civics domains), providing additional support for the notion that this particular Big Five personality factor is more about "plodding" kinds of behaviors than "dedicated" kinds of behaviors. The correlations between the Extroversion factor and knowledge scale scores are uniformly negative. This negative association between Extroversion and knowledge was contrary to our earlier predictions (Ackerman & Heggestad, 1997; however, see Gold & Arbuckle, 1990, for a review).

Multiple Regression/Correlation Analyses

Bivariate correlations between the families of predictor variables and the knowledge-scale scores provide illuminating but potentially incomplete information about the predictive validity of traits for individual differences in knowledge across domains. Questions about the incremental predictive validity of trait measures such as ability, personality, and interests can be addressed only from a multivariate perspective. Most important to some concerns, for example, is whether fluid and crystallized abilities add significantly to the prediction of.

Table 8
Personality-Knowledge Correlations

Knowledge test	Neuroticism	Extroversion	Openness	Agreeableness	Conscien.	TIE
Humanities						
American Literature	.107	-.186*	**.341****	-.078	-.217*	**.341****
Art	.065	-.196	**.371****	-.065	-.104	.298**
Geography	.066	-.234*	.090	-.042	-.270**	.170
Music	.090	-.208*	.239*	.072	-.116	.191*
World Literature	.073	-.213*	**.405****	-.003	-.190*	**.417****
Sciences						
Biology	-.112	-.162	.130	.025	-.090	.226*
Business/Management	-.026	-.041	.106	.104	.001	.119
Chemistry	-.095	-.226**	.005	.011	-.090	.109
Economics	-.009	-.207*	-.116	-.053	-.129	.107
Physics	-.070	-.166	.048	-.060	-.100	.106
Psychology	.047	-.082	.096	.084	-.043	.194*
Statistics	-.111	-.083	.076	-.050	-.138	.193*
Technology	-.040	-.249*	.229*	.071	-.122	.291**
Civics						
American Government	.025	-.225*	-.053	-.026	-.202*	.131
American History	.011	-.178	.049	-.000	-.171	.256**
Law	.048	-.138	-.072	-.022	-.066	.184*
Western Civilization	-.017	-.148	.178*	.040	-.288**	.279**
Mechanical						
Astronomy	.052	-.250*	.168	-.197	-.198	.238*
Electronics	-.166	-.068	.123	.035	-.079	.202*
Tools/Shop	-.109	-.165	.077	.060	-.125	.041

Note. Salient correlations (> .300) are shown in bold. Conscien. = conscientiousness; TIE = Typical Intellectual Engagement scale.
*$p < .05$. **$p < .01$.

knowledge after age and level of education have been accounted for. The approach that we have taken thus first enters a demographic variable (Step 1 = age) to account for individual differences in knowledge, followed by level of education (Step 2). After the variance associated with age and education, we focused on ability predictors (Step 3 = Gf estimate [a composite of math and spatial abilities]), and subsequent to Gf, the question was whether Gc added significantly to the prediction (Step 4 = Gc [a composite of the verbal abilities]). Subsequent steps focused on whether selected personality scales (Step 5 = Openness and TIE) and interest scales (Step 6 = Investigative and Artistic interests) provided any incremental prediction of knowledge-scale performance, again after all of the other variables had first been entered into the prediction equation.

The results of these analyses are presented in Table 9. For each knowledge scale, the incremental proportion of variance (R^2 to add) accounted for at each step of the analysis is shown, along with a cumulative R^2. Not at all surprising was the predictive validity of the age variable (Step 1) because this result was identical to the raw bivariate correlation. Age was a significant predictor of only Tools/Shop knowledge. Level of education (Step 2) was a significant predictor of most of the knowledge scales. It is perhaps more informative to note the seven scales for which education was not a significant predictor (Art, Geography, Psychology, Technology, Astronomy, Electronics, and Tools/Shop). Similarly, even after age and education were entered into the prediction of knowledge, the Gf composite (Step 3) was not predictive of scores on the American Literature and World Literature tests but did show incremental prediction of the other measures. The Gc composite (Step 4) showed incremental predictive validity for all measures except for the Economics and Tools/Shop tests. Such results are clearly supportive of two points: (a) Abilities are important predictors of knowledge beyond the influence that education provides, and (b) even after education and Gf are accounted for, Gc makes a significant contribution to predicting knowledge in most of the domains sampled.

The PPIK approach specifies particular personality and interest traits as being related to individual differences in knowledge. To minimize the number of variables entered into the equation, only those traits were entered into the final equation (Openness, TIE, Investigative and Artistic interests). Adding the selected set of personality variables (Step 5) and interest variables (Step 6) revealed that the previously discussed bivariate common variance could not be entirely accounted for by common variance among these and the other predictors. The personality variables in particular provided incremental prediction of knowledge scores on the American Literature, Art, World Literature, Economics, American History, and Law tests. The interest measures provided incremental predictive validity for the Art, Biology, Chemistry, and Physics tests.

Age accounted for an average of one percent of the variance in knowledge-scale scores, Education accounted for an average of 6.3 percent of additional

variance, Gf for an additional average of 13.2 percent, Gc for 8.9 percent, personality for 2.8 percent, and interests for an additional 2.4 percent of variance. In the final step for each knowledge scale, the various predictors cumulatively accounted for an average of 34.6 percent of variance, from a low of 19 percent for Business/Management knowledge to a high of 46 percent for Chemistry knowledge. Note that the order of entry of the variables ultimately determined which variable was assigned any common valid variance. Should we have entered Gc before Gf, or personality before ability, the relative contributions of the respective variables would have been different.

Summary

As predicted, although the MA group had significantly lower mean scores on numerical and spatial ability tests, they performed better than the YA norming sample on verbal ability tests. In contrast, the MA group performed significantly better than the YA group on all of the knowledge tests but one (Chemistry). Even as the traditional numerical and spatial abilities measures indicated negative effects associated with middle age, the adults in this sample knew a lot more than the YA group across a broad range of domains, including those domains that are represented both inside and outside of standard collegiate curricula.

The MA sample showed much higher scores on the knowledge scales in comparison with the YA norming sample — however, not even one significant negative correlation was found between age and knowledge-scale performance for the MA sample. One positive correlation was found for age and knowledge, namely for Tools/Shop knowledge test.

Vocational interests are very much in line with predictions from Ackerman's (1996) PPIK theory and from the review of interest-intelligence associations. Realistic, Investigative, and Artistic interests were positively associated with knowledge-scale scores across a wide variety of scales (with predictable correspondences, such as Artistic interests and Art knowledge, and Realistic interests and Electronics and Tools/Shop knowledge). Interests in Social, Enterprising, and Conventional domains were largely uncorrelated with performance on the knowledge scales. The personality trait of Openness and TIE scores were substantially positively correlated with many knowledge scales, especially in the humanities domain. Also, Extroversion was negatively correlated with performance on several knowledge scales. The final analysis consisting of multiple regressions showed clearly that fluid and crystallized abilities, personality, and interests are significant common and independent predictors of individual differences in knowledge-scale scores, even when age and educational differences among participants are first accounted for.

Discussion and Conclusions

At this stage of investigation, it is probably too soon to reach broad conclusions about the nature of adult intellect-as-knowledge, in the context of age, ability, and nonability traits. Nonetheless, the results presented here allow several observations to be made.

First of all, what predicts individual differences in knowledge? General intelligence (g) and verbal ability (even after general intelligence was partialled from g) were positively and consistently related to individual differences on the 20 knowledge scales. No consistent patterns of correlations between spatial and numerical abilities were found once general intelligence was partialled from these abilities. Clearly, the verbal tests associated with traditional assessment of crystallized intelligence are most highly predictive of standing on the knowledge scales. Note, however, that verbal ability does not account for all of the knowledge-scale performance — indeed some knowledge scales only weakly loaded on an independent Verbal Ability factor; thus, an asymmetry exists between ability measures and knowledge scales. That is, at least qualitatively, the results show that knowledge is something more than g and/or verbal abilities (or Gc as traditionally measured).

Perhaps it makes sense to return to the question of whether middle-aged adults are more intelligent than young adults. At least for the sample of participants and tests we administered in this study, the answer is that it depends. A Gf composite (mainly numerical and spatial abilities) would provide the answer that our current sample is, on average, less intelligent than the YA norming sample. A traditional Gc composite would give, as most prior research has shown, an opposite conclusion. A g composite of all three equally weighted content abilities (verbal, spatial, and numerical), still yields significantly lower performance by the current sample in comparison with the norming group. A test of the difference between the two general ability composites indicated that the current sample was significantly less intelligent than the YA norming sample, $t(276) = -3.01$, $p < .01$. A composite of knowledge tests provides an answer that is relatively unambiguous — the MA adults are on average much more intelligent than the YA group. In any composite weighting across all of these measures that does not give overwhelming emphasis to Gf measures, the MA adults would be considered on average more intelligent. Moreover, the fact that none of the knowledge tests showed significant negative age-performance correlations argues for the stability of intelligence over a significant period of middle-adult ages. Perhaps Henmon (1921) expressed it best when he stated that "intelligence, then, involves two factors — the capacity for knowledge and knowledge possessed" (p. 195).

Table 9
Summary of Hierarchical Regressions for Predicting Knowledge Scale Scores

Knowledge scale	Step 1: Age	Step 2: Education	Step 3: Gf	Step 4: Gc	Step 5: Openness, TIE	Step 6: Investigative, Artistic
Humanities						
American Literature						
R^2 to add	.017	.125**	.004	.147**	.058**	.009
Total R^2	.017	.143**	.147**	.294**	.353**	.362**
Art						
R^2 to add	.008	.032	.049*	.152**	.068*	.059*
Total R^2	.008	.040	.090*	.241**	.216**	.222**
Geography						
R^2 to add	.004	.007	.167**	.033*	.005	.006
Total R^2	.004	.10	.178**	.211**	.216**	.222**
Music						
R^2 to add	.017	.079**	.036*	.135**	.017	.001
Total R^2	.017	.097**	.133**	.268**	.248**	.286**
World Literature						
R^2 to add	.002	.095**	.021	.190**	.090**	.016
Total R^2	.017	.0.96**	.118**	.308**	.398**	.414**
Sciences						
Biology						
R^2 to add	.011	.035*	.164**	.164**	.003	.037*
Total R^2	.011	.046	.210**	.374**	.376**	.414**

Business/Management						
R^2 to add	.000	.068**	.060**	.042*	.001	.019
Total R^2	.000	.068*	.127**	.170**	.171**	.190**
Chemistry						
R^2 to add	.015	.064**	.227**	.048**	.005	.101**
Total R^2	.015	.079**	.306**	.354**	.359**	.460**
Economics						
R^2 to add	.006	.098**	.102**	.008	.052*	.013
Total R^2	.006	.104**	.207**	.214**	.267**	.279**
Physics						
R^2 to add	.003	.120**	.240**	.075**	.004	.040*
Total R^2	.003	.123**	.363**	.438**	.443**	.482**
Psychology						
R^2 to add	.000	.029	.044*	.123**	.010	.030
Total R^2	.000	.029	.073*	.196**	.206**	.236**
Statistics						
R^2 to add	.012	.110**	.099**	.031*	.007	.029
Total R^2	.012	.122**	.221**	.252**	.259**	.288**
Technology						
R^2 to add	.001	.027	.201**	.193**	.016	.026
Total R^2	.001	.028	.229**	.422**	.438**	.464**
Civics						
American Government						
R^2 to add	.001	.079**	.151**	.057**	.033	.014
Total R^2	.001	.080**	.231**	.288**	.321**	.335**
American History						
R^2 to add	.008	.153**	.083**	.059**	.048*	.001
Total R^2	.008	.161**	.244**	.303**	.351**	.352**
Law						
R^2 to add	.000	.048*	.130**	.040*	.068**	.015
Total R^2	.000	.048*	.179**	.219**	.286**	.302**
Western Civilization						

Table 9 (continued)

Knowledge scale	Step 1: Age	Step 2: Education	Step 3: Gf	Step 4: Gc	Step 5: Openness, TIE	Step 6: Investigative, Artistic
Mechanical						
Astronomy						
R^2 to add	.008	.067**	.061**	.160**	.022	.001
Total R^2	.008	.075*	.136**	.296**	.318**	.319**
(continued on next page)						
R^2 to add	.025	.021	.150**	.073**	.021	.024
Total R^2	.025	.046	.196**	.269**	.290**	.314**
Electronics						
R^2 to add	.020	.004	.333**	.031*	.020	.021
Total R^2	.20	.024	.357**	.388**	.409**	.430**
Tools/Shop						
R^2 to add	.037*	.000	.316**	.011	.012	.027
Total R^2	.037*	.037	.353**	.363**	.376**	.403**

Note. Degrees of freedom varied from one test to another because of missing data. Steps 1-4 were tests with a single degree of freedom, and the final two steps each used 2 degrees of freedom. Thus, the final-step prediction always had 8 degrees of freedom in the F ratio numerator and had an average of 106 degrees of freedom in the F ratio denomiator. Gf = fluid intelligence (a composite of math and spatial abilities); Gc = crystallized intelligence (a composite of the verbal abilities); TIE = the Typical Intellectual Engagement scale; R^2 to add = the incremental proportion of variance accounted for at each step of the analysis. *p < .05. **p < .01.

References

Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 1-27). Norwood, NJ: Ablex.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22,* 229-259.

Ackerman, P. L. (1997). Personality, self-concept, interests, and intelligence: Which construct doesn't fit? *Journal of Personality,* 65(2), 171-204.

Ackerman, P. L. (1998). Adult intelligence: Sketch of a theory and applications to learning and education. In M. C. Smith & T. Pourchot (Eds.), *Adult learning and development: Perspectives from educational psychology* (pp. 143-156). Mahwah, NJ:, Erlbaum.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121,* 219-245.

Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology,* 78, 413-432.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1,* 270-304.

Ackerman, P. L., & Rolfhus, E. L. (1998). *Knowledge structures and adult intellectual development* (College Board Research Report No. 98-3). New York: The College Board.

Baltes, P. B., & Schaie, K. W. (1976). On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. *American Psychologist,* 31, 720-725.

Baltes, P. B., & Staudinger, U. M. (1993). The search for a psychology of wisdom. *Current Directions in Psychological Science,* 2(3), 75-80.

Bayley, N. (1955). On the growth of intelligence. *American Psychologist,* 10, 805-818.

Binet, A., & Simon, T. (1916). *The development of intelligence in children.* (E. S. Kite, Trans.). Baltimore, MD: Williams & Wilkins. (Original work published 1908)

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin,* 38, 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153-193.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. *Advances in Child Development and Behavior,* 20, 91-142.

Cornelius, S. W. (1990). Aging and everyday cognitive abilities. In T. M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 411-459). Amsterdam: Elsevier.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five-Factor Inventory professional manual.* Odessa, FL: Psychological Assessment Resources. 329

Demming, J. A., & Pressey, S. L. (1957). Tests "indigenous" to the adult and older years. *Journal of Counseling Psychology,* 4, 144-148.

Dwyer, P. S. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika,* 2, 173-178.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review,* 700, 363-406.

Gilbert, J. G. (1935). Mental efficiency in senescence. *Archives of Psychology,* 27(60, Whole No. 188).

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessing typical intellectual engagement. *Journal of Educational Psychology,* 84, 537-552.

Gold, D. P., & Arbuckle, T. Y. (1990). Interactions between personality and cognition and their implications for theories of aging. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness and intervention (pp.* 351-377). New York: Elsevier.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hakstian, A. R., & Cattell, R. B. (1976). *Manual for the Comprehensive Ability Battery (CAB).* Champaign, IL: Institute for Personality and Ability Testing.

Hebb, D. O. (1941). Clinical evidence concerning the nature of normal adult test performance. *Psychological Bulletin, 38,* 593.
Hebb, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society, 85,* 275-292.
Henmon, V. A. C. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12,* 195-198.
Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6(1),* 35-45.
Holland, J. L. (1973). *Making vocational choices: A theory of careers.* Englewood Cliffs, NJ: Prentice Hall.
Horn J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Ann Arbor, MI: University Microfilms International.
Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 61-116). New York: Freeman.
Horn, J. L., & Cattell, R. B. (1966). Age differences in primary mental ability factors. *Journal of Gerontology, 21,* 210-220.
Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills and self-regulation for learning: A trait perspective. *Learning and Individual Differences, 8,* 185-209.
Lamb, R. R., & Prediger, D. J. (1981). *The Unisex Edition of the ACT Interest Inventory.* Iowa City, IA: American College Testing.
Matthews, G., & Deary, I. J. (1998). *Personality traits.* New York: Cambridge University Press.
McEvoy, G. M., & Cascio, W. F. (1989). Cumulative evidence of the relationships between employee age and job performance. *Journal of Applied Psychology, 74,* 11-17.
Miles, C. C. (1934). Influence of speed and age on intelligence scores of adults. *Journal of General Psychology, 10,* 208-210.
Mosier, C. I. (1938). A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika, 3,* 297-299.
Rolfhus, E. L. (1998). *Assessing individual differences in knowledge: Knowledge structures and traits.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.
Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88,* 174-188.
Rolfhus, E. L., & Ackerman, P. L. (2000). Assessing individual differences in knowledge: Knowledge structures and traits. *Journal of Educational Psychology.*
Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle Longitudinal Study.* New York: Cambridge University Press.
Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin, 70,* 671-680.
Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61.
Simonton, D. K. (1988). *Scientific genius: A psychology of science.* New York: Cambridge University Press.
Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research (pp.* 13-59). New York: Freeman.
Sorenson, H. (1933). Mental ability over a wide range of adult ages. *Journal of Applied Psychology, 17,* 729-741.
Sorenson, H. (1938). *Adult abilities: A study of university extension students.* Minneapolis: University of Minnesota Press.
Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31,* 811-826.
Thurstone, T. G. (1962). *PMA.* Chicago: Science Research Associates.
Tucker, L. R., & Finkbeiner, C. T. (1981). *Transformation of factors by artificial personal probability distributions* (Report No. RR-81-58). Princeton, NJ: Educational Testing Service.
Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in realworld pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology, 49,* 436-458.

The Mensa Awards for Excellence in Research

Mensa Awards for Excellence in Research are given each year to eight to ten scientists who have published outstanding research papers in peer-reviewed professional journals. This worldwide competition is sponsored jointly by the Mensa Education and Research Foundation and Mensa International, Ltd.

Typically, half the awards are given to established, senior scientists and half to researchers who have, in the past five years, entered into research into the nature of human intelligence or giftedness, education for the intellectually gifted, etc. Eligible fields of research have included psychology, education, sociology, neurology, physiology, biochemistry, and psychometrics.

Each award consists of $500 and a certificate. Many of the winning articles are reprinted in the *Mensa Research Journal.*

Judging is done by the joint American Mensa, Ltd./Mensa Education and Research Foundation Research Review Committee.

For additional information about how to enter a paper into this competition, write to MERF, Awards for Excellence in Research, 1229 Corporate Drive West, Arlington, TX 76006, USA.

Commemorate a Person or Event
Buy a Brick in the Walk of the Minds

The Mensa Education & Research Foundation offres opportunities to Mensa members to participate in the Walk of the Minds, a path of bricks commemorating special events and people.

The brick path graces the front entrance of American Mensa's National Office in Arlington, Texas.

For more information on how to purchase a brick, visit the MERF Web site at merf.us.mensa.org (click on Donations, then Walk of the Minds). You may also contact Patty Wood at the National Office, 817-607-0060, ext. 111, or email her at PattyW@AmericanMensa. org.

MERF is a 501(c)3 organization and contributions are tax deductible as specified by the IRS.

MENSA
Research Journal
Vol. 32, No. 2

Mensa Education & Research Foundation
1229 Corporate Drive West
Arlington, TX 76006-6103

*********AUTO**ALL FOR ADC 220
G#100058523 G
ERIC CLEARINGHOUSE S4 P5
1110 N GLEBE RD STE 300
ARLINGTON VA 22201-4795

97

NOTICE

REPRODUCTION BASIS

☑ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").